ESSAYS ON RACIAL ANIMUS

Mark Westcott

Dissertation
Munich, 2017

Tag der mündlichen Prüfung: 18. Januar 2018
Namen der Berichterstatter: Prof. Davide Cantoni, Prof. Joachim Winter, Prof. Florian Englmaier

ESSAYS ON RACIAL ANIMUS

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2017

vorgelegt von
Mark Westcott

Referent: Prof. Davide Cantoni, Ph.D.
Korreferent: Prof. Dr. Joachim Winter
Promotionsabschlussberatung: 31. Januar 2018

ACKNOWLEDGEMENTS

There were times—staring at regression tables, photographing seemingly endless piles of documents or programming R—that this doctorate felt like a solitary endavour. In fact, I've been helped along the way by a huge number of people. I will inevitably fail to thank them all, but there is no reason to let perfect be the enemy of good.

I have been incredibly fortunate to have Davide Cantoni as my supervisor and advisor. He has been unflinchlingly encouraging and generous with his time and advice ever since we first met. I am glad that we finally worked on a paper together. Thank you to Joachim Winter and Florian Englmaier for agreeing to serve as my second and third examiners respectively and for the comments they have provided on my research during my time in Munich.

I have much enjoyed working with my other coauthors, Felix Hagemeister, Avital Livny, and David Schindler. Avital and I have spent more hours on Skype than I care to count. Every time, I have been amazed by her enthusiasm and energy, and I am sorry that our joint project has not made it into this thesis. David Schindler and I should have spent more time in the Max-Emanuel beer garden together—the idea for our joint work was born there under the chestnut trees, each with a *Maßkrug* in hand.

Parts of this thesis were written during a research visit to the Stanford University Economics Department. Thank you to Ran Abramitzky for the invitation to visit Stanford; it was a real privilege and honour to be your student for six months. I learnt a huge amount about how to carry out, write, and present economic research. Whilst in the Bay Area, Noam Yuchtmann encouraged and cajoled me over third-wave coffee and Vicky Fouka provided much valued advice. Thanks also to David Yang, Thomas Ginn and Juan Ross for their warm welcome to a rainy Palo Alto.

The first two years of my doctoral studies were funded by the Elite Network of Bavaria through the Evidence-Based Economics programme. The first year coursework taught me a huge amount about how to do applied economics research and provided me the opportunity to attend a number of excellent lectures. As well as covering living

expenses, the programme's generous travel budget made it possible to present and discuss my work at various conferences and seminars. Thank you to Florian and Joachim again for their enthusiasm in setting up and running the programme, and to Julia Zimmermann for her patience and assistance with all matters administrative. Also at the LMU, I am grateful for a grant from the *Forschungsfonds* that enabled me to spend two weeks in Washington D.C. collecting data from the National Archives there.

Cathrin Mohr and Daniel Wissmann have been constant sources of counsel, companionship, inspiration and fun at *Lehrstuhl Cantoni*; I will miss being colleagues, but hope than we remain the good friends we have become. Thank you also to Anastasia Driva and Florentin Krämer, who dragged me away from the fourth floor bubble of Ludwigstraße 33 from time to time.

Finally, I thank Marie for believing in me and my parents Carol and Simon for their unconditional love, support and sacrifice. I am saddened that my grandparents are not alive to see the completion of this thesis; I hope it would have made all of them proud.

CONTENTS

INTRODUCTION

"Naw, Jem, I think there's just one kind of folks. Folks."

– Harper Lee, To Kill a Mockingbird

This thesis consists of three essays, all concerned with expressions of prejudicial attitudes towards racial minorities. Prejudice can have first-order welfare consequences, manifesting itself variously along a scale defined by Allport (1954): from antilocution, avoidance and discrimination through to physical attack and extermination. The causes and consequences of prejudice have been major objects of study across the social sciences. Becker (1957) introduced the topic of discrimination to the economics literature, showing that consumers' or employers' desires to avoid associating with minorities can generate wage differentials between majority and minority groups. In the Beckerian world, such anti-minority preferences are considered as exogenously determined. However, given their consequences, it is clearly also of importance to understand where such preferences come from and how they might be changed. Sociologists have long understood individuals' racial preferences to be endogenous to their environments, but the economics literature on the topic is relatively new. In a major contribution, Voigtländer and Voth (2012) show the persistence over a period of 600 years of regional variations in German antisemitic attitudes.

The first two essays of this thesis similarly link expressions of anti-minority sentiment to historic factors, providing evidence on both continuity and change in racial preferences. The first essay, which is joint work with David Schindler, examines the impact of African American soldiers in the U.K. on Britons' racial attitudes. Between 1942 and 1945 around one hundred thousand black G.I.s were stationed at military bases around the country, bringing many locals into proximity with non-whites for the first time. Using survey data covering both black G.I.s and the native population we show that these two groups interacted, resulting in a warming of attitudes on both sides. This is consistent with Allport's (1954) 'contact hypothesis', the idea that prejudice can be reduced by contact between minority and majority groups.

We test whether any such changes in attitudes might persist by making use of geographic variation in the number of black troops posted and the length of time they were posted for. Firstly, we provide evidence that the geographic distribution of black troops at any given time reflected the requirements of the U.S. military and was made without regard to any existing variation in attitudes. We do so based on documentary evidence and by showing that the distribution was orthogonal to a comprehensive set of pre-existing characteristics. The setting therefore provides quasi-experimental variation in the potential for contact between the local, white, population and black soldiers. Exploiting this, we find that areas of the U.K. where more African American soldiers were posted during World War II exhibit differential attitudes to minorities into this century. In such areas, fewer residents choose to join the British National Party, a far-right party with xenophobic policy positions. In addition, residents have less implicit (subconscious) anti-black bias, as measured by a computerised test, and are more likely to report warm feelings towards black people.

The essay bridges a large sociological literature on the contact hypothesis to an economics literature on the persistence of beliefs and preferences. Such persistence arises out of the cultural transmission of values, a model of which is provided by (Bisin and Verdier, 2001). In the model parents, motivated by imperfect altruism, make potentially costly investments in their childrens' preferences, resulting in 'vertical' (intergenerational) transmission of values. The model also allows for the 'horizontal' transmission of preferences from the wider environment to individuals, socialisation. These forces might be competing, but there are conditions under which heterogeneity in preferences can persist in a population across generations. Consistent with the theoretical model, a large number of papers have shown persistence of cultural traits, beliefs, some across extremely long time periods (these include Giuliano, 2007; Guiso, Sapienza, and Zingales, 2016; Jha, 2013; Nunn and Wantchekon, 2011).

In the second essay, which is joint work with Davide Cantoni and Felix Hagemeister, we add to this literature by examining the role of history in the recent rise of the far-right in Germany. We find that municipalities with high vote shares for the NSDAP (Nazi) party in the late 1920s/early 1930s had also higher vote shares for the newly xenophobic 'Alternative für Deutschland' (AfD) party in the 2016/17 state elections. The correlation is robust to controlling for a number of socio-economic control variables and for state fixed effects. The relationship does not hold in 2013, when the AfD ran for election on a platform of fiscal conservatism rather than xenophobia: a change which we quantify by carrying out analysis on the words used in the speeches, manifestos, Facebook and Twitter posts of all major political parties.

By examining political outcomes, the first two chapters avoid a problem that makes it difficult to identify *tastes* from observed economic outcomes. In many settings, differential outcomes between minority and non-minority groups can be rationalised

both by taste-based models of discrimination (Becker, 1957) and models of statistical discrimination (Aigner and Cain, 1977). For instance, correspondence studies have revealed that hiring managers tend to prefer white job applicants over black (Bertrand and Mullainathan, 2004, is a classic reference). Such actions might indeed be the result of prejudicial attitudes on behalf of the employer, or her expectation that customers would rather not interact with a black employee. However, in a world where productivity is not observable and minorities are on *average* less productive than whites, the hiring manager's decision could also be explained as a rational means to maximise expected productivity. When this is the case discriminatory outcomes, which have been demonstrated not only in labour markets but in goods markets too Doleac and Stein (2013), are not necessarily informative about preferences.

In the third essay, I examine an economic environment which, unusually, does allow direct inference about tastes. Concretely, I quantify the differential in tips received by white and non-white drivers of yellow cabs in New York city. At their essence, tips represent unconditional altruistic transfers from passengers, who are overwhelmingly white, to drivers. If passengers are differently altruistic towards drivers of different racial backgrounds, this should be reflected in differential tip giving. However, using an administrative dataset of over one hundred million rides, I find that non-white drivers of yellow cabs receive tips which are just one to two percent lower than those of white drivers. The differential does not become significantly larger when conditioning on a comprehensive set of journey characteristics, nor on a number of variables designed to measure variation in the quality of service provided by drivers. As such, in an environment where consumers are free to discriminate as they wish, few choose to do so—evidence that distaste of minorities is not be universal. The setting also allows me to investigate how external events effect preferences, as revealed through tips. As an illustration of the method, I show that the April 2013 bombing of the Boston marathon, an exogenously timed event which reinforced stereotypes about outgroups, did nothing to change tipping behaviour. I interpret this as further evidence that most passengers do not respond to race when making their tipping decisions.

The three essays in this thesis are self contained, and each is followed by appendices with additional material. A consolidated bibliography is presented at the end of the thesis.

1

SHOCKING RACIAL ATTITUDES: BLACK G.I.s IN THE U.K.

1.1 Introduction

Are prejudicial attitudes towards minority groups a constant of the human condition? They are certainly widespread, and a recent empirical literature has shown that anti-minority prejudice persists over the very long run, a consequence of intergenerational transmission of preferences (Voigtländer and Voth, 2012; Acharya, Blackwell, and Sen, 2016). Less is known about what it takes to change such attitudes, and whether any such changes in attitudes might themselves persist. This is important to understanding whether the consequences of prejudice—which include social conflict, hate crime, labour and goods market discrimination—are permanent or temporary features of society.

In this paper we show that the temporary presence of African American G.I.s[1] in the United Kingdom during World War II persistently reduced anti-minority prejudice amongst the British population. As the base of the U.S. military's European operation, the U.K. played host to over one and a half million U.S. troops during World War II. Around 150,000 of these troops were black, serving in segregated units with non-combat support duties such as transport and supply.[2] Both black and white G.I.s came into contact with the local population whilst off base: "Got any gum chum?" reportedly became a popular refrain amongst British children and troops of all types were frequently to be found in pubs, dance halls and restaurants (Millgate, 2010).

1. The term 'G.I.' took on its present meaning referring to military personnel during the 1940s.

2. The U.S. Army remained racially segregated until 1948. Until this point, only a very small number of black units had combat roles, most famously the aerial units which trained at Tuskegee, and the 761st Tank Battalion, the 'Black Panthers'.

Many Britons thereby saw and interacted with non-whites for the very first time. Despite pervasive racist attitudes before the war, we show evidence from surveys that these interactions were positive experiences for both the local population and for black G.I.s.

We test whether these interactions caused persistent changes in attitudes using a newly constructed dataset of U.S. military bases in the U.K. The dataset is based on station lists produced by the Office of the Adjutant General which list the units present at each military base at monthly intervals. Using this dataset, we are able to measure the racial composition of troops at around 2,000 bases.

Combining our new dataset with present-day measures of anti-minority preferences, we show that individuals in areas of the U.K. where more black troops were posted are more tolerant towards minorities sixty years after the last troops left. Firstly, we show that such areas contain fewer members of the British National Party, a far-right political party with racist policy positions. Next, we show that there is less implicit anti-black bias in these areas, as measured by Implicit Association Test (IAT) scores from Project Implicit, a website that allows interested parties to test their implicit anti-black bias. Finally, we show that those living in locations where black G.I.s were posted report warmer feelings towards black people, also using data from Project Implicit.

We explore heterogeneous treatment effects and show that persistence has occurred primarily in rural areas and areas which have not seen subsequent waves of non-white migration. This suggests that local persistence reflects intergenerational transmission combined with a lack of geographic mobility and lack of opportunities to further update beliefs.

We address two potential challenges to our identification strategy. First, an alternative explanation for our findings is that black troops were posted in areas that were already more tolerant towards non-whites, and that this tolerance has persisted until today. In fact, U.S. military policy was to place troops on the basis of military needs (Rash, 1942), and there is no evidence of any exercise to ascertain local attitudes before allocating troops. Consistent with this we demonstrate that the racial composition of troops is orthogonal to a raft of pre-existing economic, social, political, and geographic controls. Importantly, black troops are no more or less likely to be posted in areas containing a British Union of Fascists branch or areas with an existing non-white population. Second, one could also be concerned that black units were actively reallocated away from areas where they were subject to racial abuse. In fact, those race-based conflicts which did occur were between black and white G.I.s, not between black G.I.s and the local population (Smith, 1987, Chp. 6). The evidence is that the racial composition of troops at bases is orthogonal to pre-existing conditions, and that the relationship between presence of black G.I.s and tolerance is causal.

Our findings contribute to a large literature on the persistence of cultural norms, within which individual preferences are seen as endogenous to social and family environments. In the model of preference formation provided by Bisin and Verdier (2001), parents take costly investments in their children's preferences, resulting in 'vertical' transmission of values, but children are also socialised by the wider society in which they grow up, resulting in 'horizontal' transmission. Several recent empirical papers have now documented very long-run persistence in preferences, including attitudes towards minorities. Voigtländer and Voth (2012) show that variation in antisemitic attitudes in German towns and cities persisted over a time span of almost 600 years: individuals in locations which saw persecution of Jews during the middle ages were more likely to engage in antisemitic behaviour immediately prior to and during World War II. Acharya, Blackwell, and Sen (2016) show that whites living in U.S. counties with a history of slavery harbour colder feelings towards African Americans, amongst other outcomes. This is a result of institutions and norms preserved by whites in order to entrench control over African Americans after the Civil War. In contrast, our setting shows that a short 'treatment' period can result in changes in attitudes which persist through time, without institutional support or incentives sustaining them.

Our results also add to a large social science literature on the contact hypothesis (Allport, 1954). This posits that contact with minorities can reduce prejudice by causing the majority group to understand the 'essential similarity' of individuals belonging to minority groups.[3] Consistent with this hypothesis, Boisjoly et al. (2006) and Carell, Hoekstra, and West (2016) find that randomly assigning non-white roommates to white students at higher education establishments has positive effects on white students' attitudes and behaviour towards non-whites. We add to these studies by showing that changes in attitudes resulting from the treatment can persist over long time periods. In comparison to assignment of roommates at higher education establishments, our natural experiment affected a much broader cross-section of the population. Finally, whilst random roommate assignment mechanically leads to interactions, our experiment can be thought of as an 'intention-to-treat' analysis, where proximity provides opportunities for interaction, akin to situational proximity of minorities resulting from migration.

By considering effects of the presence of black troops on support for the British National Party, we also contribute to a growing literature on historical determinants of support for far-right parties. Vlachos (2017) shows that conscription of former-French citizens into the Wehrmacht during World War II permanently increased

3. This hypothesis has been a major object of study in social psychology. See Pettigrew and Tropp (2006) for a review of articles, most of which either are small-scale field or laboratory experiments or show correlations between contact and attitudes without demonstrating a causal relationship.

support for radical far-right parties, a result of political alienation. Ochsner and Roesel (2016) examine the effects of an inflow of Nazi supporters into areas of Upper Austria, showing persistent effects on support for far-right parties. These and our findings are consistent with evidence from Avdeenko and Siedler (2017), who show a high level of intergenerational correlation in attitudes towards migration and support for far-right parties.

Finally we are, to the best of our knowledge, the first to show effects of an historical event on implicit attitudes as measured by a computerised Implicit Association Test (IAT). Implicit attitudes are described as "traces of past experience [that] affect some performance, even though the influential earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection" (Greenwald and Banaji, 1995, p. 4f.). Implicit attitudes against minority-groups are increasingly used in the economics literature to measure bias (e.g. Lowes et al., 2015), and have been shown to be predictive of behaviour in a number of domains, including in hiring decisions (see Uhlmann et al., 2009, for a review). However little has been shown so far about their determinants, malleability and persistence. Our research shows that implicit attitudes can be shaped by historical events, suggesting they are subject to the same kinds of forces of transmission as stated and revealed preferences.

The paper proceeds as follows: Section 1.2 provides further historical background, Section 1.3 provides evidence on the contact which occurred between black troops and the local population. Section 1.4 describes the data, identification strategy and other necessary preliminaries for the data analysis. Section 1.5 documents the effect of historical inter-group contact on support for the British National Party, and Section 1.6 provides evidence from other outcome measures. Section 1.7 concludes.

1.2 Historical Overview

The United States entered World War II in December 1941 following a declaration of war from Germany and its immediate reciprocation by the U.S. Congress. Headquarters for the U.S. military's European operation were established in London the next month and the first combat troops soon arrived via ports in Northern Ireland. The first aerial bombardments of Germany were carried out in June out of bases in Norfolk, part of the Eastern Base Section.[4]

In addition to hosting the European headquarters and providing a base for aerial operations the U.K. also functioned as a staging and training post for the ground troops who would later liberate France and eventually Germany. These troops began

4. The Eastern Base Section, one of four areas created in order to decentralize operations, was predominantly used by the Army Air Force (Waddell, 2010, p. 142). See Appendix Figure A1 for a map of base sections.

arriving in the U.K. in May 1942 in preparation for a mid-1943 land offensive. This planned operation, codenamed 'Bolero', was later cancelled and as a result troop numbers declined from a peak of 230,000 to around 100,000 as troops were reallocated to North Africa or the Pacific. The build-up of troops began again in May 1943 once plans for a 1944 offensive were settled. By November, around 160,000 troops were arriving per month. Troop numbers reached their peak in June 1944 with one and a half million G.I.s stationed in the U.K. Figure 1.1 shows the numbers of troops in U.K. for each month between 1942 and 1945.

Troops were stationed throughout the country, in rural and urban areas alike. The influx of troops into a small country put huge strain on available accommodation; troops were stationed wherever space could be found.[5] Troops were mainly accommodated in newly constructed camps or ex-RAF or British Army quarters. After 1943, the supply of such accommodation was exhausted and some troops were billeted in private homes. To the best of our knowledge black troops were never accommodated this way.

Most American ground troops left England in the course of 'Operation Overlord', the airborne and amphibious assault of occupied Europe beginning on 6th June 1944. On the first day of the operation alone 150,000 troops landed in Northern France via beaches in Normandy. By the time the operation ended in August 1944, just 700,000 G.I.s were left in Britain, down from the June peak of one and a half million. Units continued to cross to Europe, but troop numbers in the U.K. did not decrease much further until the end of the war: the U.K. continued to serve as the headquarters of operations in Europe, as a base for Army Air Force units, as the point of entry for American troops bound for continental Europe and as the main location of military hospitals in Europe. However, by November 1945, almost all American units had left the U.K.

1.2.1 Black G.I.s

Over 900,000 African Americans served in the U.S. military during World War II, more than half of them outside of the U.S. (Moore, 2013). As in previous wars, black soldiers served in racially segregated units, normally under command of white officers. With few exceptions, black troops were limited to non-combat 'labour' or 'service' roles, most often supply and quartermaster services, transport, food preparation and sanitation.[6]

5. Scotland was deemed unsuitable for military facilities, presumably because of its distance from mainland Europe: only very few troops were ever stationed there (Lee, 1966, p. 623).

6. Racism was institutionalised in the military. General Patton's views expressed in a letter to his wife, "A colored soldier cannot think fast enough to fight in armor." (Sasser, 2014, p. 104), seem to be representative of military leaders' attitudes at the time.

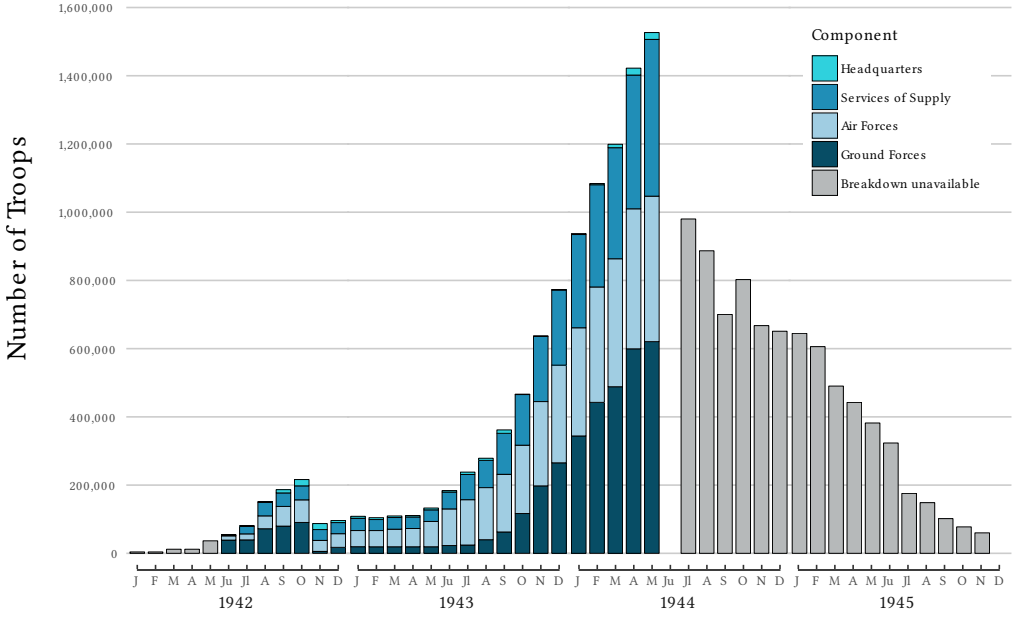


Figure 1.1. Build up of U.S. Army strength in the U.K from January 1942 to November 1945 and, where available, a breakdown according to type. Data for June 1944 are unavailable. Sources: Ruppenthal (1978, p. 232) and Pogue (1954, p. 541).

U.S. Army Enlistment Records, which provide a record of all individuals who served in the Army during World War II, reveal that the black G.I.s were diverse with regards to their demographics. Appendix Figure A2 shows that around 50% of African Americans serving in the Army had no high school education, a relatively moderate positive selection with respect to the young black male population as a whole, where around 70% of 18 to 30 year-olds had no high school education.[7] Black G.I.s were drawn from both northern and southern states: Appendix Figure A3 shows a map of the states of birth of African Americans who served in the military during World War II.

Many black G.I.s served abroad; around ten percent of G.I.s who served in the U.K were African American.[8] The British government initially requested that the U.S. military refrain from sending black troops to Britain. Ostensibly this was to avoid causing conflict between white G.I.s and British citizens, who might show "more effusiveness to the coloured people than the Americans would readily understand"[9], but a desire to minimize the non-white presence in the U.K. and concerns about sexual activity no doubt also played a role. The matter was taken up by the Foreign Secretary Anthony Eden, who claimed that the British climate was 'badly suited to negroes', and likely also by Churchill himself (Reynolds, 2006). Nonetheless, the request was refused, both for practical reasons—black troops made up a considerable share of the support units which were of military necessity—and out of U.S. political concerns. Suggestions were made to limit black troops to port areas, where the U.K.'s small existing black population was concentrated, but these were rejected. The policy instead was to "place them [black troops] where needed" (Rash, 1942).[10]

The U.S. army's colour bar was maintained; interaction between black and white soldiers was minimised, with accommodation, dining and training facilities all segregated. A 'pass system' was introduced in order to keep black and white troops apart during their leisure time, with black and white units allowed off base on different days of the week or assigned different venues to visit.

---

7. Data on the population of black males is provided by the 1940 Census. The slight positive selection on education reflects actions of local draft boards and army requirements on literacy/education. Until 1941, draft boards had discretion to decide whether a potential recruit had the mental capacity to serve; in this period, 'mental deficiency' was the most frequent cause for a pre-inductee to be rejected for service by a draft board. In 1941, a fourth-grade level literacy requirement was introduced, followed by the Army General Classification Test, introduced in 1943. Although the later test purportedly measured generalised intelligence, it has been argued that it in reality measured educational attainment, which was on average lower amongst the black population (Murray, 1971).

8. These appear to have been slightly positively selected in terms of education compared to the universe of black G.I.s, see again Appendix Figure A2.

9. E. Bridges to J. Martin, 21 July 1942, in Franklin D. Roosevelt Library, Hyde Park, New York, U.S.A, Harry Hopkins papers, box 136.

10. Units arriving in Britain were first assigned to one of several armies by the European Theater Headquarters, and would then be assigned to a base by their army's headquarters.

The U.K. government was at pains to take no overt actions to enforce segregation, refusing for example to instruct police officers to recommend segregation to local bar and restaurant owners. There are however a few isolated examples of local authorities attempting to limit contact between black G.I.s and British women (Reynolds, 2006, p. 123). Nonetheless, evidence suggests that frequent contact between soldiers and local populations took place: troops visited local bars, restaurants and dance halls during their leisure time.[11] There are also frequent reports of younger British women visiting black troops on base (Smith, 1987, p. 187 ff).

1.3 Evidence on Contact

In most areas of the where black G.I.s were stationed, locals would have been seeing and interacting with black people for the first time.[12] Despite evidence of wide-spread racial prejudice before the war[13], existing evidence suggests that the British responded positively to black G.I.s, whilst attitudes towards white soldiers were more mixed. Writing in a biweekly newspaper, George Orwell remarked that "the general consensus of opinion seems to be that the only American soldiers with decent manners are the Negroes" (Orwell, 1943). This attitude is also on display in words attributed to an unknown Englishman in 1943, "I don't mind the Yanks, but I don't care much for the white fellows they've brought with them" (Olson, 2010, p. 287) and in a British woman

11. Officially black and white units had the same quota for leisure time, although some black soldiers complained about white officers restricting their passes (Smith, 1987, p. 134).

12. There is no administrative information on the size of the black population the U.K. before the introduction of an ethnicity question in the 1991 census. General Eisenhower wrote to General Lee in 1942: "There is practically no coloured population in the British Isles." (September 5 1942, ETOUSA AG 291.2-B, available online at `https://archive.org/stream/IndoctrinationOfPersonnelArrivingInTheUK/IndoctrinationOfPersonnelArrivingInTheUK_djvu.txt`). The permanent Asian and black population of the British Isles was estimated in 1939 at around 7,000 individuals (Cabinet Office "Report of the Working Party on Coloured People Seeking Employment in the United Kingdom", 17/12/1953 in CAB124/1191, via Little (1998)). Existing evidence suggests that this population was concentrated around port cities such as London, Liverpool, South Shields and Cardiff (Spencer, 1997). In a later section, we use surnames and places of birth from the full count 1911 census to estimate the geographic distribution of non-white migrants in the U.K.

13. The best source on racial attitudes in Britain before World War II is Little (1998), who provides a narrative of changes in attitudes to non-whites from the 17th century onwards. In a unique study, Lapiere (1928) surreptitiously questioned 315 individuals in London, Birmingham, North Wales and Liverpool, asking each of them a variant of the question "Would you let children associate with those of good coloured people?". Based on their responses, respondents were classified as being 'without prejudice' (4 percent), 'doubtful cases' (15 percent) or 'with prejudice' (81 percent), without much variation according to social class. In addition, twenty hotels were contacted and asked "Does the management permit either African or Indian guests?". Only four hotels responded affirmatively, although the author points out that at one such London hotel, admission was limited to Indian nobility. Lapiere's results from England compare unfavourably to a similar study he carried out in France, where the majority of individuals questioned reported no colour prejudice.

writing to a friend "Everybody here adores the Negro troops, all the girls go to their dances, but nobody likes the white Americans. They swagger about as if they were the only people fighting this war" (Reynolds, 1995, p. 303).

Below we add to this anecdotal evidence using surveys of U.S. troops stationed in the U.K. and qualitative responses to a questionnaire of U.K. citizens carried out in 1944.

1.3.1 U.S. Military Surveys

Two surveys carried out by a research branch of the U.S. War Department provide evidence on troops' contact with and attitudes towards the English. These surveys are amongst over one hundred carried out during the war in order to provide army command with information about the attitudes of soldiers (Stouffer et al., 1949). Sampling for each survey took place in two stages; in the first stage, army units were selected from data on troop locations. Sampling in this stage is described as 'not strictly random', since the War Department could reduce costs by sampling multiple units at the same base. However, the aim was for the sample to be representative of the population of units in terms of branch and type of unit. In the second stage of sampling, men were sampled from their units by random selection from a duty roster.

The surveys useful for our purposes are "Attitudes Towards The British" (S-122) and "Attitudes Toward Army Life" (S-92). The first of these, carried out in December 1943, surveyed 3,261 individuals with the aim of understanding soldiers' attitudes towards the British, and provides evidence on the amount and mode of contact between troops and the British population[14]. The survey asks how many local people of each of a number of categories—families, men in the armed forces, girls in the armed forces, civilian girls, civilians (older men and women)—the respondent has got to know 'fairly well', in each case being asked to select between none, one or two, several, or a very large number. 86 percent of troops reply that they know at least some families or civilians fairly well. The most common way of meeting English[15] people is described as 'a chance meeting' (67 percent), with lower numbers being introduced through a friend (18 percent) or through a service organisation or some other way (14 percent). The most common meeting places were eating places or pubs (33 percent), around town (33 percent) and at dances (17 percent). The number of civilians known to soldiers increases with time in the U.K. (see Appendix Figure A4), consistent with interactions between troops and civilians taking place on an impromptu basis.

In Survey S-92, carried out in November 1943, troops stationed in Britain were asked "How has your opinion of the English people changed from what it was before you came to England?" and "How do you think the English people's opinion of Americans

14. Although the survey provides some demographic data (education level and marital status), race was not recorded.

15. English is likely a misnomer for British.

has been changed by having American soldiers in England?". Unlike the December survey, this survey reports whether respondents are black (422) or white (2,257). We visualise the responses to these questions separately for black and white troops in Figures 1.2 and 1.3. The median white G.I. has reduced his opinion of the English since being stationed and believes that the English have lower opinions of Americans as a result of the presence of G.I.s. Strikingly, the pattern for black G.I.s is very different: the majority of black G.I.s have positively updated their view of the English whilst being stationed and simultaneously believe that the British have positively updated their opinion of the Americans.[16] We interpret this as evidence of 'high-quality' interactions between black G.I.s and the local population, leading both black G.I.s and the local population to update their beliefs about each other.[17]

1.3.2 Mass Observation

We provide more qualitative evidence on British attitudes to black troops, and the changes caused by contact with them, using data from Mass Observation. This was a U.K. based survey organisation founded in 1937 aiming to create an 'anthropology of ourselves' by regularly collecting written testimony from a panel of volunteer respondents around the country (Madge and Harrisson, 1937). This panel was not designed to be representative of the population, and the occupational listings of participants shows that the panel consists largely of middle- and upper-class professionals. The panellists were asked to keep personal diaries which they sent to Mass Observation on a monthly basis. In addition, they responded to monthly 'directives' from Mass Observation which aimed to collect information on opinions about a wide variety of topics. In June 1943, the panel was asked 'What is your personal attitude towards coloured people, and is there any difference in your attitude towards members of different coloured races? Have wartime events or experiences had any effect on your attitudes in this respect?'. The question was listed as 'Priority B', with the instruction that respondents should answer it if they have time to do so.[18] We collected and digitised all responses to the question[19], of which thirty three make explicit mention of African American soldiers. These responses are reproduced in full in Section A.4.

16. Regressions reported in Appendix Table A1 show that the difference in responses given by white and black troops are statistically significant and robust to controlling for the base at which individuals are stationed, their branch of the army, state of birth, rank and education levels.

17. In doing so, we assume some reflection; that is, when a soldier is asked how he thinks that English opinions of Americans have changed, his response reflects how English opinions of individuals like him have changed.

18. In this month, Priority A questions dealt with BBC news bulletins and attitudes towards dentists and dentistry.

19. The Mass Observation archive is in the care of the University of Sussex. We accessed photos of the archive in September 2015 via Mass Observation Online, `http://www.massobservation.amdigital.co.uk/`

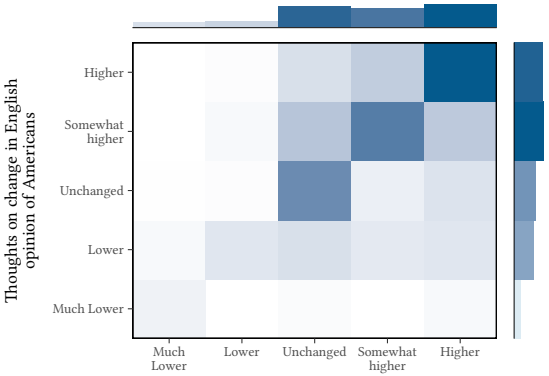


Figure 1.2. Density plot of individual black G.I.s' responses to the questions 'Has your opinion of the English people changed from what it was before you came to England?' (horizontal axis) and 'Do you think English people's opinion of Americans has been changed by having American soldiers in England' (vertical axis). Darker cells indicate more mass. The sample is 442 black G.I.s posted in Britain in November 1943.

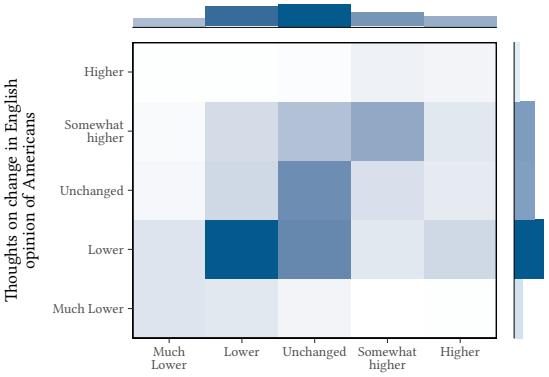


Figure 1.3. Density plot of individual white G.I.s' responses to the questions 'Has your opinion of the English people changed from what it was before you came to England?' (horizontal axis) and 'Do you think English people's opinion of Americans has been changed by having American soldiers in England' (vertical axis). Darker cells indicate more mass. The sample is 2,257 white G.I.s posted in Britain in November 1943.

We manually classify these responses along two dimensions – sentiment towards black troops (positive or negative) and whether the respondent has interacted with black troops. Eight responses mention or imply contact with African American soldiers, of which six (75%) show positive sentiment towards them, whilst a narrow majority of respondents who make no mention of contact show negative sentiment. This might be a result of selection into contact rather than changes in attitudes caused by such contact—however, six responses explicitly report positive updating of attitudes or beliefs based on the presence of black troops, with only two implying negative updating. Examples of positive updating are "My little contact with the American Negroes made me more sympathetic to them. They liked being over here, because they were treated better here by us than by the white Americans in their own country" and "Have wartime events and experiences had any effect on my attitude[?] The answer is yes. The presence of many more American negroes in this country may make me take less interest them through accepting them as normal and familiar". These kinds of responses seem to reflect exactly the kinds of changes in attitudes through contact that Allport's contact hypothesis postulates.

1.4 Estimation Framework

Having provided suggestive evidence of changes in contemporary attitudes caused by the presence of black G.I.s, we now test for persistent effects of their presence. Concretely, we use two independent data sets with high geographical resolution to test whether proximity to black G.I.s persistently affected local racial attitudes. The first of these provides data on membership of the British National Party, a far-right xenophobic party, across all of England and Wales' 180,000 census output areas ('neighbourhoods'), the lowest level geography on which the national statistics agency collects data. The second is an extract from the data generated by 'Project Implicit', a website on which individuals can carry out a test for their implicit racial attitudes. As well as carrying out the test, individuals are prompted to provide self-reported racial attitudes and demographic information including their postcode.

We describe the datasets in more detail later, but first lay down the shared estimation strategy that we apply to all outcome measures.

1.4.1 Troop Data

Our empirical exercise exploits a new dataset of U.S. army units in the U.K.[20] The dataset is based on the station lists produced by the U.S. Army Adjutant General's Office (AGO) on a monthly basis throughout the war. These station lists, which list all units stationed in the United Kingdom and their location at a snapshot in time, were produced by the (AGO) out of daily 'morning reports' produced by each military

20. Full descriptions and sources for all datasets used in the paper are given in Section A.2.

unit. These daily morning reports were punched onto IBM cards by Mobile Records Units and then transported to the Adjutant General's Office, who consolidated them in order to create station lists, which were distributed amongst military commanders.

Most station lists for the period June 1943-1953 survive and are housed at the U.S. National Archives in Washington D.C. We were kindly provided with digitisations of sixteen station lists by Captain Philip Grinton and we digitised a further eleven ourselves, resulting in twenty-seven digitised station lists. These lists cover months from June 1943 to December 1945 and are listed in Appendix Table A2. Appendix Figure A5 shows an extract from one of these lists.

Each line in a station list represents a unit, which is identified with an abbreviation of its name (e.g. 1944 QM TRK CO for the 1944th Quartermaster Truck Company), and is listed along with the coordinates of the base at which the unit was posted at[21], the nearest town or village, and a symbol to indicate if the unit is a segregated unit with African American troops.

Using these lists, we are able to create a dataset of the units posted at twenty-seven points in time at 1,937 military bases/camps (unique according to their coordinates). These bases were located widely across England and Wales, with the notable exception of the South East of England. This was a consequence of the plan for the invasion of occupied France: U.S. ships departed from the west of England, the British Navy from the east, and vessels landed on French beaches in the same formation. Appendix Figure A6 shows a map of all bases/camps in the dataset.

1.4.2 Identification

Our aim is to estimate the effects of the local presence of black units on racial attitudes. To do so we use variation across space in both the time black troops were posted for and their numbers.[22]

Recall that the vast majority of black units were 'support units', providing services such as transport and sanitation. In our preferred specification, we control for the presence of all support units, both black and white. Again, we do so taking into account both the number of units posted and the time they were present for. This

21. Map coordinates are provided in a coordinate system defined on the Cassini projection, for example WL5715 for a base in Watford. The first two digits of the coordinates indicate an 100 by 100km square, the subsequent digits provide the northing and easting from the bottom left of that square, to an accuracy of 1 kilometer. We reproject coordinates to the British National Grid using a Cassini projection with false easting 500000, false northing 100000, central meridian -1.19276, scale factor 1.0 and latitude of origin 50.617708.

22. If interactions between troops and local population happened spontaneously, as the evidence in Section 1.3 suggested, the probability of locals coming into contact with troops increases in the number of troops stationed nearby and the length of time they were stationed for.

ensures we are not simply just capturing variation in the local presence of support units, but rather exploiting variation across bases in the racial composition of support units.[23]

We begin by identifying support units based on their names; the list of unit names indicating support status is shown in Appendix Table A3. This list covers 91% of units described by the station lists as being a segregated black unit.[24] Then we count the number of black support units posted at each base in each month we have data for, $BlackSupportUnits_{b,m}$ where b is a base and m a month in which we observe the allocation of troops. Next, we sum across all such months to produce a base-level measure, $BlackUnitMonths_b$:

$$BlackUnitMonths_b = \sum_m BlackSupportUnits_{b,m}$$

We create an analogous measure for the presence of support troops:

$$SupportUnitMonths_b = \sum_m SupportUnits_{b,m}$$

Our identification assumption is that the presence of black troops at a base (measured by $BlackUnitMonths_b$), is exogenous to pre-existing racial attitudes in the population around that base, controlling for the overall presence of support troops (measured by $SupportUnitMonths_b$). Our estimation results would be biased if, when deciding which support units to allocate to which base, black units were strategically stationed in areas with particular racial attitudes. There is no evidence for this being the case. As discussed in the introduction, there were some early suggestions to limit black troops to port areas (where Britain's small existing black population was concentrated), but these were rapidly dismissed, as indicated by a 1942 internal memo from the Services of Supply Headquarters, "The policy has been defined to place them [black units] where needed"; that is, military constraints determined the allocation of troops to bases.

In addition, we now show that, after controlling for the degree of presence of support units, the degree of presence of black units at a base is uncorrelated with a large

23. In an alternative specification, unreported, we instead exploit variation in the intensity of presence of black units given the intensity of presence of *all* military units, without any substantive changes to results. However, given that the vast majority of black units were carrying out support rules, we want to make sure that our results are not being driven by differences in the types of locations that support and combat troops were posted to.

24. Units described in the records as being black but not classified as support units include 'Detachment of Patients' (i.e. those receiving care at a military hospital), 'Detachment of Prisoners', and the few combat black units which served in the U.K.: the 320rd Anti-Aircraft Barrage Balloon, the 333rd Field Artillery Battalion, the 969th Artillery Battalion, the 452nd Anti-Aircraft Artillery Battalion, the 578th Field Artillery Battalion, the 614th Tank Destroyer Battalion, and the 999 Field Artillery Battalion. We ignore these units in our analysis.

number of pre-existing economic, political and geographic characteristics around that base. For each of these characteristics, we estimate the following regression equation:

$$BlackUnitMonths_b = \alpha + \beta X_b + \gamma \cdot SupportUnitMonths_b + e_b$$

where b is a military base, X_b is the variable of interest and e_b the error term. In order to account for potential correlation in the error term between observations, we cluster standard errors at the modern local authority level, which divides England and Wales into 348 administrative regions. Results are shown in Table 1.1, which each line corresponds to a regression on a separate pre-existing characteristic; full descriptions of the variables are provided in Section A.3. Depending on the data source, the characteristics vary either at the parish level, the local-government district level, the constituency level or, for some geographical measures, are directly measured at the coordinates of the base.

The first group of variables shown in Table 1.1 measure economic conditions around the base; they include the parish-level population density as measured in the 1931 census, the parish-level rate of population growth between 1921 and 1931, sectoral composition of employment at the local government district level, unemployment rate at the local government district level and a binary indicator as to whether the base was inside an urban area.

The second group of variables measure political and societal features. Firstly, we use the full count of the 1911 England and Wales census to estimate the number of non-white migrants in each parish as per 1911.[25]. Our most direct measure for pre-existing racial attitudes is constituency level data on the presence of a branch of the British Union of Fascists (BUF). The BUF was formed in 1932 by Oswald Mosley, a British politician who had served as a Member of Parliament for both the Conservative and Labour parties. Xenophobia was 'a mainstay of the rhetoric of the BUF' (Redvaldsen, 2016).[26] Internal Labour Party research in 1934 aimed to ascertain whether a branch had been formed in each of England and Wales' constituencies, data which we collect

25. The construction of this variable, which uses place of births and names to infer non-white migrant status, is outlined in Section A.3 A map of its distribution is shown in Appendix Figure A7. There is no administrative information on the size or distribution of the non-white population before the introduction of an ethnicity question in the 1991 census. 1911 is the last year for which full count census data has been released.

26. The focus was clearly antisemitic, although there is evidence of more general racism; a party speech by William Joyce in 1934 talked about removing foreigners "be they Hebrew or any other form of alien", whilst a party pamphlet contained the text "Under Fascism, no alien shall enter this country to take the jobs of Britons, and aliens who are already here who have abused the hospitality of this nation will be sent back whence they came", ("Fascism explained: Ten points of fascist policy", accessed as LP/FA3/34/542 at the People's History Museum Archive, University of Central Lancashire).

and code.[27] Finally in this group, we collect data on the share of the electorate voting for the Conservative party in the last pre-war election, that in 1935.

The third and final group of variables are various measures of geographic isolation. We calculate the distance from each base to the nearest local government district with urban status in 1931, to the nearest town/city with a 1939 population over 100,000 (of which there are 55), and to the nearest city with a 1939 population over 300,000 (of which there are 8). We collect data on the location railway stations in 1939 in order to calculate the distance from the base to the nearest railway station, and digitise a 1946 road map in order to calculate the distance from each base to the nearest major road.

Apart from one variable, the indicator for being inside an urban local government district (significant at the 8% level), we find no statistically significant correlations between pre-existing characteristics around bases and the presence of black troops there. This pattern is consistent with the idea that troop placements were made on the basis of *military* requirements, orthogonal to local conditions, and as such any correlation between the presence of black troops and contemporary anti-minority attitudes reflects a causal effect.

1.4.3 Treatment Definition

Our hypothesis is that, in areas around military bases with black troops, interactions led to changes in local attitudes, and that these attitudes have since persisted, a result of intergenerational transmission. The question remains of how to match contemporary populations to historic bases. In our main specifications, we consider a contemporary location to be 'treated' by a given military base if the location and the military base share a common postcode district.[28] That is, we consider any base in the same postcode district as a contemporary location to have contributed to historical contact. Our treatment is defined on the postcode district level as follows:

$$BlackUnitMonths_j = \sum_b \sum_m \mathbb{1}[b \in j] \cdot BlackUnits_{b,m}$$

where j indexes postcode districts, $\mathbb{1}[b \in j]$ is an indicator function for whether base b is inside postcode j and $BlackUnits_{b,m}$ is the number of black support units posted

27. Data is held and was accessed as LP/FAS/34 at the People's History Museum/University of Central Lancashire.

28. Postcode districts cover very small areas in cities but larger areas in more rural areas (65% of the variance in log-area is explained by log-population density), with a median area of 27 square kilometres. We use postcode districts, with their varying sizes, to define our treatment since this captures the logic that contact between troops and individuals a fixed distance apart was more likely in less densely populated areas. Our results are not sensitive to this choice: Appendix Table A4 shows regressions where treatment is defined at the local government district level instead).

Table 1.1: Effect of covariates on the presence of black troops

	$\hat{\beta}$	p-value $H_0 : \beta = 0$
Economic:		
Population density in parish, 1931	−0.006	0.84
Rate of population growth in parish 1921-1931	0.029	0.25
Agricultural employment in local government district, 1931	−0.018	0.51
Professional employment in local government district, 1931	−0.039	0.17
Unemployment rate in local government district, 1931	0.044	0.12
Urban local government district, 1931	0.049	0.08
Social and political:		
Non-white migrants in parish, 1911	0.000	1.00
Conservative vote share in constituency, 1935	−0.027	0.35
British Union of Facists branch in constituency, 1934	−0.007	0.85
Geographic:		
Distance to coast	−0.034	0.24
Distance to nearest city, 1939	0.006	0.86
Distance to nearest large city, 1939	−0.015	0.62
Distance to nearest urban district, 1931	−0.010	0.80
Distance to railway station, 1939	0.046	0.32
Distance to major road, 1939	−0.006	0.85

Notes: Each row reports beta coefficients from a regression of the form $BlackUnitMonths_b = \alpha + \beta X_i + \gamma \cdot SupportUnitMonths_b + e_b$, where X_b is a control variable and e_b is an error term. The sample is all bases hosting at least one support unit at one point in time.

at base b in month m. We create an analogous measure for the presence of support troops:

$$SupportUnitMonths_j = \sum_b \sum_m \mathbb{1}[b \in j] \cdot SupportUnits_{b,m}$$

These measures capture the intensity of presence of black and support troops around a given contemporary location. Because we want to exploit variation in the number of support units around a given location which were black, our population of interest consists of individuals living in postcode districts where at least some support units were located. To illustrate the variation that we exploit, Figure 1.4 shows how the ratio of $BlackUnitMonths_j$ to $SupportUnitMonths_j$ varies across postcode districts.[29]

1.5 Long-Term Effects on British National Party Membership

We now estimate the effects of proximity to black troops on contemporary attitudes towards minorities. We begin by measuring local attitudes using data on local membership of the British National Party (BNP), a far-right political party with extreme positions on race.[30]

The British National Party was founded in 1982 as a splinter group from the National Front, an openly racist organization with links to European neo-Nazis. The party's founder was jailed for conspiracy to incite racial hatred in 1986, and a senior official described the party as '100 per cent racist' in 1995 (BBC, 2001). Police officers and prison officials are banned by their employers from joining the party, and in 2009 a government minister attempted to introduce a similar ban for teachers, citing a desire to 'keep racism ...out of our schools' (The Guardian, 2009).

Attempting to increase its electoral relevance, the party began to outwardly reject claims of racism in 1999, but its ideology is still widely considered to be just that (see e.g. The Spectator, 2009). Most tellingly, non-white members were banned from the party until this was deemed illegal by a court in 2010. Even whilst espousing a 'modernization' agenda, Nick Griffin, the party's leader from 1999 to 2014 continued to call for the 'repatriation' of non-white Britons, who a party manual referred to as

29. Appendix Figure A8 and Appendix Figure A9 show maps of the two variables separately.

30. This variable measures preferences at the tail of the distribution, but the evidence suggests that this correlated with preferences more generally. Biggs and Knauss (2011) show that, on the constituency level, BNP membership is related to the share of the population voting for the BNP in the 2005 national election, a less extreme measure. Table A5 shows, again at the constituency-level, that BNP membership is correlated with unease about the cultural effects of migration, even once controlling for the unemployment rate and non-white population share in the constituency.

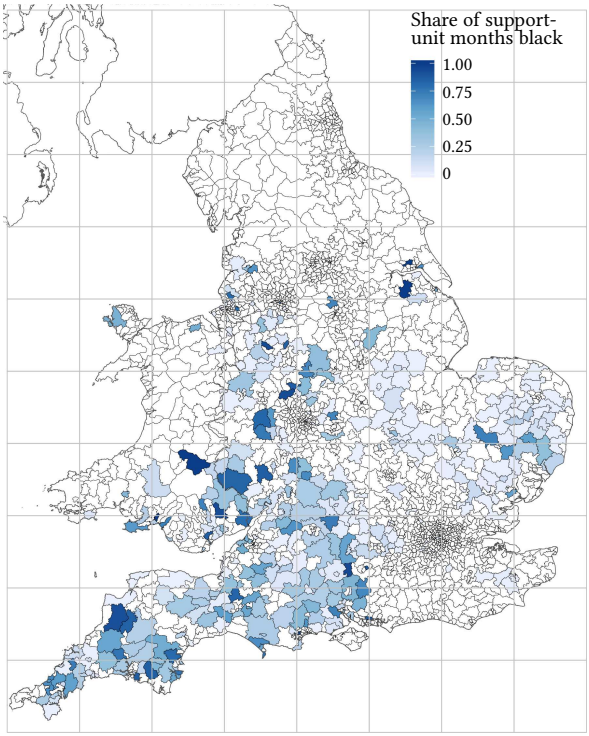


Figure 1.4. Figure shows variation across English and Welsh postcode districts in the share of support-unit months which are due to black units. Also shown are grid-cells used to generate grid-cell fixed effects

'racial foreigners'. Further evidence of the party's racist ideology is provided by the BNP's website, extracts from which from 2008 are displayed in Section A.1.

In addition, we provide suggestive evidence of the opinions of British National Party members using a survey carried out by YouGov on behalf of Goodwin and Evans (2012), who kindly shared their data. YouGov contacted 2,951 members of its national panel who, according to their data, had indicated being supporters of the British National Party, the United Kingdom Independence Party or the English Defence League. The response rate was 73%, including 54 members and 58 former members of the BNP. The opinions of these individuals on race/migration are reported in Figure 1.5. A large majority believes that immigrants are the main cause of crime and of disease, and 47 percent believe in innate differences in intelligence between black and white Britons. Majorities also believe in the unconditional repatriation of foreigners and reject the concept of non-white Britishness. Although we can not guarantee that the respondents to the survey are a true random sample of BNP members, the evidence suggests that members have extreme views on race.

Our regressions make use of data from Biggs and Knauss (2011), who geolocate members members of the party using a membership list published online in 2008.[31] The list was confirmed by the party to be genuine and is understood to provide a complete listing of members of the party in November/December 2007, although the BNP has claimed that the list contains a number of ex- and prospective members too. Membership of the British National Party was not contingent on the presence of a local branch; at the time the membership list was published, individuals were able to join the party by completing a paper or online form. Since joining the party entailed some cost[32], membership data provides a revealed preference measure of racial attitudes.

The membership list comprises information on 13,009 individuals, including a home address with a valid U.K. postcode in 97% of cases.[33] Biggs and Knauss (2011) aggregate the data to the 2001 'output area' level (the lowest level geography on which the Office of National Statistics aggregates demographic and social data) and report the number of members within each area. The authors kindly shared their dataset with us. We match the data to the most recent neighourhood definitions, resulting in data on the universe of the census output areas across England and Wales, which we refer to as neighbourhoods. Of the 184,109 neighbourhoods in England and Wales (median population: 303), 12,513 (6.7%) include at least one BNP member. The maximum

31. Due to legal constraints in Germany, where most of the analysis was carried out, we were not able to process this information directly from the original, leaked membership list.

32. Membership for pensioners, students and the unemployed cost £15 a year, a standard membership cost £30 a year and a 'gold' membership £60.

33. There are over one million unique postcodes throughout the U.K., with each postcode covering on average 15 households.

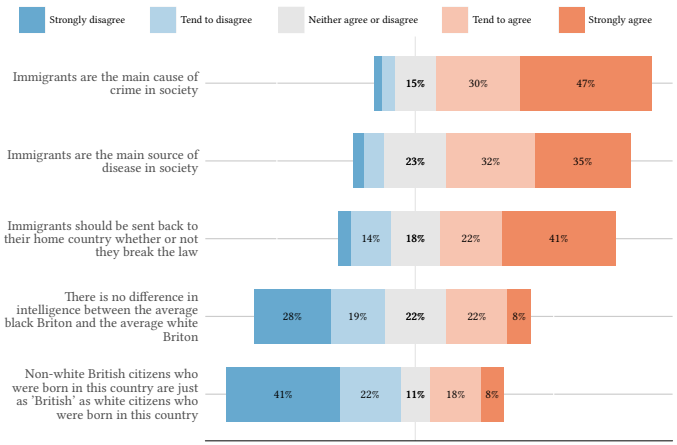


Figure 1.5. Opinions of 112 members and former members of the BNP. Source: Goodwin and Evans, 2012.

number of members per neighbourhood is 11, a neighbourhood in Barnsley, in the north of England. The neighbourhoods containing BNP members are displayed in Appendix Figure A10, which also shows the UK geographies with active BNP branches in 2008.[34] The figure reveals that membership of the BNP is geographically diverse, and members live in areas both with and without local branches.

1.5.1 Estimation and Results

Our dependent variable is the number of BNP members per 100,000 white residents in a neighbourhood.[35] We limit our sample to neighbourhoods inside England and Wales since negligible numbers of troops were posted in Scotland and because Northern Ire-

34. Data on the existence of BNP branches comes from the BNP newspaper, the 'Voice of Freedom', Issue 100. Last accessed August 2017 at `https://www.scribd.com/document/25905300/Voice-of-Freedom-101`.

35. As described above, the data on BNP membership is provided at the neighbourhood level. We run our main regression at the neighbourhood level too, without further aggregation, since this allows us to accurately match each unit to larger historical geographies. This simplifies the generation of control variables, since neighbourhoods almost always fit completely inside historic boundaries, something which would not be the case for larger geographic units. Appendix Table A6 reports results from regressions at the postcode district level, albeit without a full set of controls. These are in-line with our main specification. In addition, Section 1.5.3 reports estimation results from alternate model specifications.

land contains only a legible numbers of BNP members. In constructing our dependent variable we divide by the number of white residents in the neighbourhood to avoid a mechanical correlation between the outcome measure and the size of non-white population.[36] Further, as described in Section 1.4, we limit our sample to neighbourhoods in postcode districts where support units were posted, so as to exploit variation in the racial makeup of support troops posted around any given location in our estimation.

We thus estimate regression equations of the form:

$$\begin{aligned} \textit{BNP members per 100,000 whites}_i = \alpha + \beta_1 \textit{BlackUnitMonths}_j \\ + \beta_2 \textit{SupportUnitMonths}_j + \boldsymbol{X}_i + \ u_i \end{aligned} \tag{1.1}$$

where i is a neighbourhood, j stands for that neighbourhood's postcode district, $\textit{BlackUnitMonths}_j$ and $\textit{SupportUnitMonths}_j$ are constructed as per Section 1.4, $\boldsymbol{X}_i$ is a vector of controls and u_i is the error term. In order to account for correlation in the error term between observations, we cluster standard errors at the modern local authority level, which divides England and Wales into 348 administrative regions. We carry out randomization inference as an alternative method for accouting for this correlation in Section 1.5.3. Our key parameter of interest is β_1, the estimated effect of the presence of black troops, conditional on the presence of support troops, on BNP membership.

Results of estimating Equation 1.1 are displayed in Table 1.2, with each column representing a separate regression with a different set of control variables. To ease the interpretation of coefficients, both the explanatory and dependent variables are standardised to have zero mean and a standard deviation of one. The regression shown in Column 1 includes no additional controls. We introduce further controls in a piecewise fashion, starting with Column 2. First we include grid-cell fixed effects, generated by imposing an arbitrary eight by ten grid on the map of England and Wales and matching neighbourhood to these cells (the cells are shown in Figure 1.4). Estimation in this and subsequent columns exploits *within grid-cell variation* in the presence of black troops, comparing neighbourhoods which are relatively close to one another but which vary in terms of the number of black units posted. This is a more flexible way to capture unobserved variables with regional variation than latitude/longitude controls.

For the remaining columns, we match contemporary neighbourhoods to their historic geographies: 1931 parishes, 1931 local government districts and 1935 constituen-

36. Recall that membership of the party was limited to whites at the time the membership list was published, although given the party's policy platform it seems unlikely that many non-whites would choose to join the party.

Table 1.2: Effect on BNP membership

	BNP members per 100,000 whites (std.)				
	(1)	(2)	(3)	(4)	(5)
Black unit-months	-0.022***	-0.023***	-0.022***	-0.023***	-0.029*
	(0.008)	(0.008)	(0.008)	(0.009)	(0.01)
Support unit-months	✓	✓	✓	✓	✓
Grid-cell fixed effects		✓	✓	✓	✓
Economic controls			✓	✓	✓
Geographic controls				✓	✓
Socio-political controls					✓
Clusters	234	234	234	234	172
Observations	48,732	48,732	48,665	48,665	26,498

Notes: Each column reports coefficients and standard errors (in brackets) from an OLS regression where both the dependent variable and main independent variable have been standardised to have mean zero and standard deviation one. The unit of observation is the neighbourhood (2011 census output area). The outcome variable is BNP members per 100,000 white inhabitants. Black unit-months, the reported independent variable, is our measure of the presence of black troops in the neighbourhood's postcode district. Support-unit months is the measure of presence of support troops in the neighbourhood's postcode district. Economic controls are: population density (in the 1931 parish), change in population (in the parish between 1921 to 1931), the share of employment due to the agricultural sector (in the 1931 local government district), the share of employment due to the professional sector (in the 1931 local government district) and a dummy variable for urban status (1931 local government district). Geographic controls are distances to the coast, to the nearest large city, to the nearest city, to the nearest urban district, to the nearest railway station and to the nearest major road. Social- political controls are the estimated number of non-white migrants in the parish (1911), the share of the votes in the 1935 constituency going to the Conservative party, and a binary indicator for the presence of a British Union of Fascist branch in the constituency, measured in 1934. Standard errors are clustered at the local authority district level and reported in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

cies, using boundary data provided by the Vision of Britain Project. In most cases, neighbourhoods are completely contained within these geographies; if not, they are matched to the geography which contains the neighbourhood's population weighted centroid. Based on these historic geographies, we add controls for pre-existing economic conditions in Column 3, for geography in Column 4 and for socio-political variables in Column 5. These are the same set of controls we used to demonstrate exogeneity of the treatment measure in Section 1.4.2.

In the baseline specification, Column 1, a one standard deviation increase in the presence of black troops, as measured by $BlackUnitMonths_j$ reduces the number of BNP members as a share of white citizens by 0.02 of a standard deviation. The effect is reasonably small, as might be expected given the length of time that has passed since the treatment, but is highly statistically significant. Comparing coefficients across the columns shows that adding grid-cell fixed effects, economic controls or geographic controls makes only a marginal difference to the estimated coefficient (this follows from the exogeneity of the treatment established in Section 1.4.2). The estimate shown in Column 5, which includes socio-political controls, is estimated with less precision because we do not have data on the existence or otherwise of a British Union of Fascist or the share of Conservative party votes for all constituencies. However, the coefficient is still statistically significant at the 10% level and is in fact larger than that of the previous columns.[37]

1.5.2 Geographic decay

In this section, we investigate what level of proximity is required between a base at which black troops were posted and a contemporary location in order for the presence of troops to have had an persistent effect on attitudes. To do so, we assume that a military base affected a neighbourhood if the neighbourhood is within a given distance of the base, k. For various values of k, we generate a treatment variable which measures the presence of black units in a k-kilometre radius around the neighbourhood:

$$BlackUnitMonths_{i,k} = \sum_b \sum_m \mathbb{1}[d(i,b) \leq k] \cdot BlackUnits_{b,m}$$

where j indexes postcode districts, $d(i,b)$ is the Euclidean distance in kilometres between the population-weighted centroid of neighbourhood i and base b and $BlackUnits_{b,m}$

37. An alternative explanation for the results reported in Table 1.2 is that the presence of black troops caused individuals with anti-minority preferences to migrate away. However, the evidence is that the disruption to the housing market during war time made such moves unlikely. Appendix Figure A11 shows that the building of new houses all but ceased during the war period, and Appendix Figure A12 uses an online archive of local newspaper adverts to show that the supply of non-new properties also collapsed.

Table 1.3: Effect on BNP membership, alternative treatment measures

	Dependent variable:			
	BNP members per 100,000 whites (std.)			
	(1)	(2)	(3)	(4)
	Baseline	+ Grid-cell F.E.s	+ Economic Controls	+ Geography Controls
Panel A: binary treatment				
Black units ever stationed	-0.056***	-0.033	-0.030	-0.037
	(0.021)	(0.020)	(0.020)	(0.020)
Panel B: months				
Months with black units (std.)	-0.027***	-0.019**	-0.018**	-0.019**
	(0.009)	(0.009)	(0.009)	(0.009)
Panel C: horserace				
Black units ever stationed	-0.047*	-0.030	-0.026	-0.030
	(0.025)	(0.024)	(0.024)	(0.024)
Months with black units (std.)	-0.003	0.008	0.007	0.008
	(0.014)	(0.011)	(0.012)	(0.012)
Black unit-months (std.)	-0.013	-0.025***	-0.025**	-0.025**
	(0.011)	(0.009)	(0.010)	(0.010)
Support unit-months	✓	✓	✓	✓
Grid-cell fixed effects		✓	✓	✓
Economic controls			✓	✓
Geographic controls				✓
Clusters	234	234	234	234
Observations	48,732	48,732	48,665	48,665

Notes: Each column within each panel reports coefficients and standard errors from an OLS regression. The unit of observation is the neighbourhood (2011 census output area). The outcome variable is BNP members per 100,000 white inhabitants in the neighbourhood. The dependent variables vary by panel. In Panel A, the independent variable is a dummy for whether black units were ever stationed in the neighbourhood's postcode district. In Panel B, the independent variable is the number of months for which black units were posted (standardized to have mean zero and standard deviation one). Regressions shown in Panel C includes both of these variables, as well as our regular measure of presence of black troops, 'black unit-months'. Control variables are as per Table 1.2. Standard errors are clustered at the local authority district level. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

is the number of black support units posted at base b in month m. We create an analogous measure for the presence of support troops:

$$SupportUnitMonths_{i,k} = \sum_{b} \sum_{m} \mathbb{1}[d(i,b) \leq k] \cdot SupportUnits_{b,m}$$

We then use these new measures in regressions of the form:

$$\begin{aligned} BNP\ members\ per\ 100{,}000\ whites_{i,k} = \alpha + \beta_1 BlackUnitMonths_{i,k} \\ + \beta_2 SupportUnitMonths_{i,k} + \mathbf{X}_i + u_i \end{aligned} \tag{1.2}$$

Note that the treatment measure now varies on the neighbourhood level, the level at which we observe our outcome measure. The results of these regressions, for distances between two and ten kilometres, are visualised in Figure 1.6; all variables are again standardised, so that the coefficients can be interpreted as beta coefficients. The magnitude of the point estimate decreases with distance, and at a radius of five kilometres or above, the effect is no longer statistically significant from zero. As such, the effect of the black G.I.s appears to be highly localised. This is consistent with our assumption that the closer a location was to a base, the more likely it was that contact took place between the population there and any black G.I.s on the base.

1.5.3 Robustness

In this section we examine the extent to which the results presented in Table 1.2 are robust to changes in specification or the sample.

Firstly, in Table 1.3, we examine robustness to alternative treatment indicators. We progressively add controls in the same fashion as in Table 1.2. In the regressions reported in Panel A, we use a binary indicator as our main independent variable; this takes the value of one if black units were ever posted in a neighbourhood's postcode district and zero otherwise. Across specifications, the estimated effect of having any black troops posted in a neighbourhood is negative. However, it is only statistically significant in Column 1, where we control only for the degree of presence of support units. Introducing grid-cell fixed effects, as per Column 2 reduces the size of the estimate and its relative precision. This is likely a result of there being little residual variation in the variable once controlling for grid-cell fixed-effects.

In Panel B, we count the number of months (maximum twenty-seven) in which any black units were present in a neighbourhood's postcode district for. This measure has more variation than that used in Panel A, and remains significant and negative across

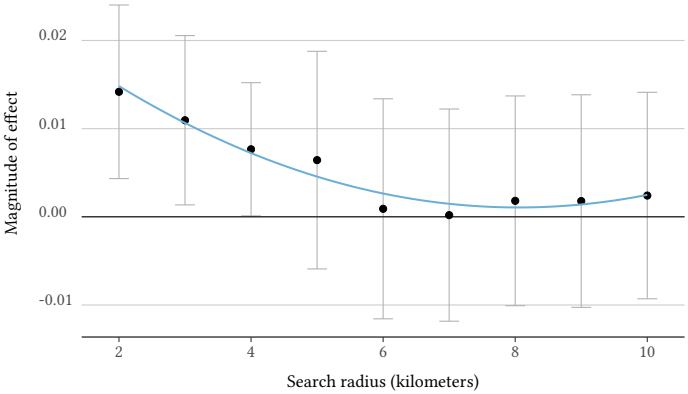


Figure 1.6. Geographic decay. Point estimates and 95% confidence intervals are displayed for eight regressions, varying the radius around a location at which bases are counted as contributing to potential contact, i.e. varying k in Equation 1.2.

all specifications; the number of BNP members in each neighbourhood reduces with the number of months black units were posted.

Finally, in Panel C, we run a 'horserace', where we include both of these alternative measures as well as our preferred measure, unit-months. In the specifications with controls, Columns 2 to 4, the coefficient on our preferred measure is significant, whilst the other measures are not. The evidence suggests that both the number of black troops and their duration of stay matter for contemporary outcomes. Our preferred treatment measure, unit-months, increases in both of these.

In Table 1.4, we show that our results are robust to changes in the specification of the model. In Column 1, we run a logistic regression on a dependent variable that indicates whether there is at least one BNP members in a given neighbourhood. Multiple BNP members within a given household or spillover effects between neighbours might increase BNP membership within a given neighbourhood, without necessarily indicating stronger racial prejudice. If this is the case, the absolute value of members in an output area is no more informative of racial prejudice than a binary variable that indicates whether members are present or not. In Column 2, we estimate a negative binomial model, which explicitly accounts for the fact that BNP members per neighbourhood is a count variable. When estimating the negative binomial regression, we constrain

Table 1.4: Effect on BNP membership, alternate models

	(1)	(2)	(3)	(4)
	Logit model	Negative binomial model	Dep. var. IHS transformed	Dep. var. IHS & Logit
Black unit-months (std.)	-0.0035***	-0.0038***	-0.012*	-0.045**
	(0.001)	(0.001)	(0.007)	(0.02)
Support unit-months	✓	✓	✓	✓
Grid-cell fixed effects	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓
Model	Logit	Neg. Bin.	OLS	Logit
Dep. var. transformation	Std.	Std.	IHS	IHS
Clusters	234	234	234	234
Observations	48,664	48,665	48,665	48,664

Notes: Coefficients from several regressions. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 100,000 white inhabitants, except for Column 2 where it is a dummy whether the neighbourhood contains any BNP members. The independent variable is our measure for potential contact with black troops 'Black unit-months'. We control for 'support unit-months' in each regression. Standard errors are clustered at the local authority district level. Standard errors in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table 1.5: Effect on BNP membership, subsamples

	Dependent variable:		
	BNP members per 100,000 whites (std.)		
	(1)	(2)	(3)
	Inland	Not London	No BNP Branch
Black unit-months (std.)	-0.034***	-0.021**	-0.022***
	(0.009)	(0.009)	(0.008)
Support unit-months	✓	✓	✓
Grid-cell fixed effects	✓	✓	✓
Economic controls	✓	✓	✓
Geographic controls	✓	✓	✓
Clusters	179	226	220
Observations	27,048	47,937	46,092

Notes: Each column reports coefficients and standard errors from an OLS regression. The unit of observation is the neighbourhood (2011 census output area). The outcome variable is BNP members per 100,000 white inhabitants in the neighbourhood. Black unit-months, the reported independent variable, is our measure for the presence of black troops in the neighbourhood's postcode district. Control variables are as per Table 1.2. The sample varies across columns. The sample for Column 1 is all neighbourhoods at least 20 kilometers away from the coast. The sample for Column 2 is all neighbourhoods apart from those inside Greater London. The sample for Column 3 excludes neighbourhoods in districts where a local branch of the British National Party is active. Standard errors are clustered at the local authority district level. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

$ln(Whites_i)$ to one to account for the fact that BNP membership in any neighbourhood is constrained by the number of whites living there. In Column 3, we transform our dependent by taking its inverse hyperbolic sine (Burbidge, Magee, and Robb, 1988). The distribution of BNP members across neighbourhoods is highly skewed, and the inverse hyperbolic sine transformation makes the distribution more normal, reducing the influence of outliers. In Column 4, we both transform the independent variable and estimate a logit model. The coefficients are not comparable across specifications, but coefficients on *BlackUnitMonths* are negative and statistically significant in all regressions. This suggests that the results found in our main specification are not driven simply by model selection.

Table 1.5 reports results on selected samples. In all cases, we use the same set of

controls as in Column 4 of Table 1.2, including geographic and economic controls. In Column 1 we drop from the sample any neighbourhoods within twenty kilometers of the coast. This excludes all locations in which the small pre-war black population of England and Wales was concentrated. Results remain highly significant. We exclude London from the sample in Column 2, and in Column 3 exclude all areas with an active BNP branch in 2007. Across all our specifications, the coefficient on *BlackUnitMonths* remains highly significant.

Further robustness checks are reported in the Appendix. In Appendix Table A7, we show our results are robust to weighting neighbourhoods by the size of the population. In Appendix Figure A13 we report results of regressions where we leave counties out of the sample one at a time, and show that the results are not driven by any particular county in the UK. Finally, in order to assuage concerns that spatial correlation might cause biased estimates of standard errors, Appendix Figure A14 reports the results of randomization inference. We randomly reassign the share of support unit-months due to black units, $BlackUnitMonths_b/SupportUnitMonths_b$, between bases in order to create a simulated assignment of $BlackUnitMonths_b$. We then estimate Equation 1.1 using the simulated $BlackUnitMonths_b$ and the real $SupportUnitMonths_b$. We repeat the exercise 8,000 times, generating a distribution of estimates of β_1, which is centred around zero. The estimate of β_1 with the real $BlackUnitMonths_b$ is larger in magnitude than all but 0.1% of the counterfactual estimates.

1.5.4 Heterogeneity in Persistence

In order to make some progress around understanding channels of persistence, we examine heterogeneous effects along two dimensions in Table 1.6. The analysis is in the style of Voigtländer and Voth (2012), who demonstrates that the persistence of anti-Semitism in Germany across a period of 600 years 'fails' in cities which are open to trade and migration.

We begin by splitting the sample by the rural-urban status of the neighbourhood.[38] Column 1 reports the results of the specification of estimating Equation 1.1 with controls as per Column 4 of Table 1.1 on the sample of neighbourhoods in rural areas. Column 2 does the same for urban areas. The results reveal that the effect of black troops on BNP membership is about double as large in rural areas than in urban areas, standardised coefficients of −0.031 compared to −0.015. In fact, the effect in urban areas is not statistically significant at any conventional level, despite a larger sample size.

38. The Office of National Statistics classifies neighbourhoods as belonging to one of four urban categories or six rural categories based on the typical settlement of the households in the neighbourhood. Urban categories are major conurbation, minor conurbation, city and town, city and town in sparse setting. Rural categories are town and fringe, town and fringe in sparse settings, village, village in spare setting, hamlets and isolated dwelling.

These results make sense given the low levels of internal migration rural areas of the U.K. where, according to data from the UK household longitudinal study 'Understanding Society'[39] 25% of individuals were living within two miles (3.2 kilometres) and 47% within five miles (8 kilometres) of the place they were grown up. Urban areas are however magnets for migration, both international and domestic. As such, in rural areas more of the population will be direct descendants of those living in the same area during World War II. Given this, a model of vertical transmission of attitudes from parents to children would predict more persistence in rural compared to urban areas, precisely what we find.

In Columns 3 and 4 we investigate heterogeneity according to non-white share. In areas with larger non-white populations, individuals have more opportunities to update their beliefs about non-whites, and we would therefore expect less persistence compared to areas which are still predominantly white. To test this hypothesis, the regression reported in Column 3 include onlys neighbourhoods which are part of a local authority area with a bottom-quartile share of non-whites, in Column 4 we only include neighbourhoods which are part of a local authority area with a top-quartile share of non-whites. We indeed find that the effect of black G.I.s is almost twice as large in areas which are predominantly white.[40]

1.6 Additional Outcome Measures

In this section, we investigate the effect of black troops on two further measures of anti-minority prejudice. The outcome measure used so far, membership of the BNP party, is appealing on account of being a 'revealed' preference measure and being something that we can measure for the population of the U.K. However, it is also, as we have shown, a measure of *extreme* racial attitudes. Whilst extreme prejudicial attitudes probably have the most impact on minority individuals' welfare (leading, e.g. to hate crime and overt discrimination), it is also of interest whether the presence of black G.I.s lead to changes in attitudes more widely construed. We make progress on this question by examining measured implicit anti-black bias and survey data.

1.6.1 Implicit Attitudes

First, we show that presence of black troops has had persistent effects on *implicit* anti-black attitudes. Implicit attitudes, "traces of past experience [that] affect some performance, even though the influential earlier experience is not remembered in the

39. University of Essex. Institute for Social and Economic Research. (2016). Understanding Society: Innovation Panel, Waves 1-8, 2008-2015. [data collection]. 7th Edition. UK Data Service. SN: 6849.

40. In unreported results, we tested for effects of the local presence of black G.I.s on the size of the non-white population, and found no statistically significant relationship. Non-white migration to the U.K. has largely been to a handful of major cities.

Table 1.6: Effect on BNP membership, heterogeneity

	Dependent variable:			
	BNP members per 100,00 whites (std.)			
			Non-white share	
	(1)	(2)	(3)	(4)
	Rural	Urban	Bottom Quartile	Top Quartile
Black unit-months (std.)	-0.031***	-0.015	-0.037**	-0.022
	(0.008)	(0.01)	(0.02)	(0.03)
Support unit-months	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓
Clusters	196	204	63	61
Observations	15,659	33,006	13,233	10,415

Notes: Each column reports coefficients and standard errors from an OLS regression. The unit of observation is the neighbourhood (2011 census output area). The outcome variable is BNP members per 100,000 white inhabitants in the neighbourhood. Black unit-months, the reported independent variable, is our measure for the presence of black troops in the neighbourhood's postcode district. Control variables are as per Table 1.2. The sample varies across columns. In Column 1 the sample is neighourhoods in rural areas, in Column 2 it is neighbourhoods in urban areas, in Column 3 it is neighbourhoods in local authority districts with a bottom quartile non-white share, in Column 4 it is neighbourhoods in local authority districts with a top quartile non-white share. Standard errors are clustered at the local authority district level. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

usual sense" (Greenwald and Banaji, 1995, p. 78), have been shown to be predictive for outcomes in a variety of domains when measured by a computerised test.[41]

We use data from Project Implicit, an American non-profit organization which hosts several country-specific websites allowing users to test their implicit attitudes using various Implicit Association Tests ('IATs'). IATs are widely used in psychology (see Greenwald, McGhee, and Schwartz, 1998, for a review), and increasingly in economics,

41. For an overview see Uhlmann et al. (2009). Amongst these, Agerström and Rooth (2011) find that hiring managers' discrimination against obese job applicants is predicted by a test that measures implicit attitudes towards the overweight. In an incentivised experiment, Stanley et al. (2011) show that white subjects' implicit racial attitudes predict their judgement about the trustworthiness of randomly matched black partners, even conditional on their reported racial attitudes.

as a way of measuring implicit attitudes (e.g. Lowes et al., 2015). The test, which is computer administered, can be applied to measure implicit associations on any topic, but the most common application has been to race. A consistent finding (at least in the US) is of a strong automatic preference towards white people, although there is considerable variation in race IAT scores between individuals. A full description of the test is give in Section A.5.

Project Implicit hosts IATs on a range of subjects (e.g. gender, sexuality, weight), but by far the most popular test is that for race. After completing the test, subjects are asked to answer several questions regarding their attitudes towards religion, minorities, politics and supply their postcode and general demographics. About half of U.K. residents taking the IAT test do so on the U.S. website. Following an IRB exemption, we received a dataset containing all data collected by the U.K. and U.S. IAT websites between 2004 and 2013.[42] The dataset from the U.K. website contains around 240,000 started sessions, 90,000 of which were seen through to completion. Of these, valid postcode data is provided in 25,826 sessions. We combine these with 19,582 valid observations from the U.S. website to complete our sample of 45,408 sessions. We make no claims about the representativeness of the sample: the median age is 29 years, compared to 40 for the population, and roughly two thirds of the participants are male. Subjects are also better educated than the population as a whole. Nonetheless, the sample contains individuals from throughout the U.K. (all 348 of England and Wales' local government districts are represented) and of a wide range of ages (13 to 89).[43]

In order to test whether contact with black G.I.s has had persistent impact on IAT scores, we follow the same basic empirical strategy as in Section 1.5, but now run individual-level regressions on the "D" score, the summary metric produced by the test, which is also standardised to have mean zero and a standard deviation of one. Higher scores indicate a stronger automatic association of positive words with white faces, with a score over zero indicating implicit racial bias toward whites (Greenwald, McGhee, and Schwartz, 1998). Results are presented in Table 1.7. In the whole sample the effect is insignificant, but the coefficient points to the presence of black troops lowering implicit bias against blacks. Once we limit the same to rural districts, estimated effect sizes are larger and statistically significant. In these areas, a one standard deviation increase in the presence of black troops associated with a 7% of a standard deviation reduction in implicit anti-black bias—around double the effect on membership of the British National Party.

42. A public version of the U.S. dataset, without zip codes, is available via the Open Science Framework at `https://osf.io/y9hiq/`

43. Curiosity seems to be the main driver for participation in the test. The most common stated reasons for coming to the website are the recommendation of a friend and mention in news articles. Around 20% of tests in the dataset were conducted in the days after the BBC publicised the test on its news website in April 2005 under the headline "Are you racist? The test that claims to know" (BBC, 2005)

Table 1.7: Effect on implicit anti-black bias

	All			Rural		
	(1)	(2)	(3)	(4)	(5)	(6)
Black unit-months (std.)	-0.0317*	-0.0270	-0.0294	-0.0655***	-0.0658**	-0.0702**
	(0.0164)	(0.0182)	(0.0185)	(0.0250)	(0.0255)	(0.0272)
Support unit-months	✓	✓	✓	✓	✓	✓
Demographic Controls		✓	✓		✓	✓
Grid fixed effects		✓	✓		✓	✓
Economic controls			✓			✓
Location controls			✓			✓
Clusters	220	220	220	156	156	156
Observations	5,468	5,430	5,420	1,435	1,427	1,425

Notes: Each column reports coefficients and standard errors (in brackets). The unit of observation is the individual. The dependent variable is the standardized measure of implicit anti-black attitudes from the IAT. The independent variable is our measure for contact with black troops 'Black unit-months' in the postcode district. Demographic controls are age, age squared and gender. Standard errors in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

1.6.2 Thermology

Using the same dataset as above, we now examine the effect of the presence of black troops on stated attitudes. After completing the IAT test, participants are asked a range of survey questions, including a number of 'thermology' questions of the form "Please rate how warm or cold you feel toward the following groups (0 = coldest feelings, 5 = neutral, 10 = warmest feelings)". We take as our dependent variable the thermology score towards black people (or, if the test is being taken on the US website, 'African Americans') and we control for the thermology score towards white people. Results are displayed in Table 1.8. We find the same pattern: a positive but insignificant relationship between warmness of feelings to blacks and black troops when considering the whole sample, but a stronger and statistically significant relationship when the sample is limited to rural districts. The effect sizes are almost identical to those we find for the IAT results, with a one standard deviation increase in the presence of black troops increasing warmth towards black people by 7.3% of a standard deviation in the specification with all controls, Column 6.

Table 1.8: Warmness of feelings towards blacks

	All			Rural		
	(1)	(2)	(3)	(4)	(5)	(6)
Black unit-months (std.)	0.0126	0.0121	0.0111	0.0511*	0.0652**	0.0732**
	(0.0160)	(0.0170)	(0.0175)	(0.0267)	(0.0319)	(0.0309)
Support unit-months	✓	✓	✓	✓	✓	✓
Demographic Controls		✓	✓		✓	✓
Grid fixed effects		✓	✓		✓	✓
Economic controls			✓			✓
Location controls			✓			✓
Clusters	220	220	220	156	156	156
Observations	5,329	5,295	5,285	1,393	1,386	1,384

Notes: Each column reports coefficients and standard errors (in brackets). The unit of observation is the individual. The dependent variable is the reported thermology score, indicating how warmly individuals feel on a scale from 0 to 10 towards black people. The independent variable is our measure for contact with black troops 'Black unit-months' in the postcode district. Demographic controls are age, age squared and gender. All regressions control for thermology scores towards white people. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

1.7 Conclusion

We have investigated the effect of the presence of black American troops in the United Kingdom during World War II on anti-minority prejudice. Large numbers of black G.I.s were posted in the U.K. at a time where the population was almost exclusively white. Documentary evidence suggests that the allocation of black troops to military bases in the U.K. took place without consideration of local racial attitudes, and we show that the allocation is orthogonal to a large set of economic, political and social variables. As such, variation across bases as to the number of black units posted allows us to identify causal effects of the local presence of troops.

We found that areas of the U.K. in which black soldiers were posted during World War II contain fewer members of the British National Party, a far-right party with racist policy positions, in 2009. The effect is particularly strong in rural areas – that is, areas where population movements are lower and which remain predominantly white. In addition, individuals in such areas exhibit less implicit anti-black bias, as measured by a computerised Implicit Association Test, and are more likely to report warmer feelings towards black people. Taken as a whole, our results provide support

for the 'contact hypothesis' (Allport, 1954), which postulates that contact between groups can reduce animosity towards the minority group, and show that such effects can persist in geographies across time.

It is interesting to note that the contact which we describe meets many of the conditions that Allport postulated were necessary for intergroup contact to lead to improved relations: equal status, common goals, intergroup cooperation and personal interaction. Black G.I.s were in the United Kingdom for a relatively short period of time, were there to support the war effort, and did not compete for jobs or public goods with the local population. More work is required to understand how the *mode* of interaction between groups affects any changes in attitudes that contact might produce.

Appendix A.1: Policies of the British National Party

"The British National Party exists to secure a future for the indigenous peoples of these islands in the North Atlantic which have been our homeland for millennia. We use the term indigenous to describe the people whose ancestors were the earliest settlers here after the last great Ice Age and which have been complemented by the historic migrations from mainland Europe" (Mission statement posted on BNP website, 2007, accessed June 2016 via `http://archive.org`)

"IMMIGRATION - time to say ENOUGH! On current demographic trends, we, the native British people, will be an ethnic minority in our own country within sixty years" (BNP Manifesto, 2007, accessed June 2016 via `http://archive.org`)

"We don't 'hate' black people, we don't 'hate' Asians, we don't oppose any ethnic group for what God made them, they have a right to their own identity as much as we do, all we want to do is to preserve the ethnic and cultural identity of the British people" (BNP website, 2007, accessed June 2016 via `http://archive.org`)

Appendix A.2: Data Sources

U.S. Army station lists

U.S. Army Station Lists are housed in the National Archives in Washington D.C. as Record Group 407, HMS Entry Number NM 3 377 A. Catalog entry at http://research.archives.gov/description/6883370.

U.S. Army enlistment records

World War II Army Enlistment Records, created, 6/1/2002 - 9/30/2002, documenting the period ca. 1938 - 1946. National Archives and Records Administration. Office of Records Services - Washington, D.C. Modern Records Programs. Electronic and Special Media Records Services Division. (1998 -) Available online at https://aad.archives.gov/aad/series-description.jsp?s=3360.

Mass Observation directive responses

Mass Observation directive responses are downloaded from Mass Observation Online, http://www.amdigital.co.uk/m-products/product/mass-observation-online.

War Department surveys

The American Soldier in World War II: Attitudes Toward Army Life, Nov, 1943 [dataset] USAMS1943-S092, Version 1. Directed by Dr. Samuel A. Stouffer for the Research Branch, Information and Education Division, War Department [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Aug-23-2017. Accessd online at https://ropercenter.cornell.edu/CFIDE/cf/action/catalog/abstract.cfm?type=&start=&id=&archno=USAMS1943-S092

The American Soldier in World War II: Attitudes in Field Forces, Dec, 1943 [dataset]. USAMS1943-S112, Version 3. Directed by Dr. Samuel A. Stouffer for the Research Branch, Information and Education Division, War Department [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Aug-23-2017. Accessed online at https://ropercenter.cornell.edu/CFIDE/cf/action/catalog/abstract.cfm?type=&start=&id=&archno=USAMS1943-S112

BNP membership

Details on BNP Membership at the 2001 Output Area level was kindly provided to us by Biggs and Knauss (2011).

Survey of current and former BNP members

Extracted from a survey carried out by YouGov on behalf of Goodwin and Evans

(2012), who kindly provided us with their data. YouGov is a polling organisation which maintains a panel of around 350,000 respondents. These were screened for individuals who had previously reported supporting, voting for or membership of the UK Independence Party, the British National Party, or the English Defence League, who were then contacted by YouGov and asked to take part in the survey.

Results from Implicit Association Tests

Data comes from Project Implicit, an American non-profit organization which hosts several country-specific websites allowing users to test their implicit attitudes using various Implicit Association Tests (Xu, Nosek, and Greenwald, 2014). Following an IRB exemption, we were provided with a dataset containing all data collected by the U.K. and U.S. IAT websites between 2004 and 2013[44].

2011 U.K. census summary

Office of National Statistics (ONS). Available online at `https://www.nomisweb.co.uk/census/2011`.

Full count England and Wales census, 1911

"England and Wales Census, 1911." [Database] FamilySearch. `http://FamilySearch.org`: 14 June 2016. From "1911 England and Wales census." Database and images. findmypast. `http://www.findmypast.com` : n.d. Citing PRO RG 14. The National Archives of the UK, Kew, Surrey.

Population of cities, 1939

Southall, H.R., Aucott, P., Dorling, D., Ell, P. (2004). Great Britain Historical Database: Census Data: Age and Sex Statistics, 1851-1971. [data collection]. UK Data Service. SN: 4551, `http://doi.org/10.5255/UKDA-SN-4551-1`

Parish and local-government statistics, 1921 and 1931

Great Britain Historical GIS Project (2017) 'Great Britain Historical GIS'. University of Portsmouth. `http://www.visionofbritain.org.uk/`

Parish and local-government boundaries, 1921 and 1931 and constituency Boundaries, 1935

Great Britain Historical GIS Project (2017) 'Great Britain Historical GIS'. University of Portsmouth. `http://www.visionofbritain.org.uk/`

44. A public version of the U.S. dataset, without zip codes, is available via the Open Science Framework at `https://osf.io/y9hiq/`

Electoral results, 1935

Field, W. (2007), British Electoral Data, 1885-1949. [data collection]. UK Data Service. SN: 5673. http://doi.org/10.5255/UKDA-SN-5673-1

Labour Party questionnaire on local fascist activities

Held and accessed as LP/FAS/34 at the People's History Museum/University of Central Lancashire.

Ordnance Survey planning maps, 1946

Available online from the National Library of Scotland at http://maps.nls.uk/view/91542574

Railway Network Inspire dataset

Published by Network Rail and available at https://data.gov.uk/dataset/railway-network-inspire. Accessed April 2016.

Appendix A.3: Construction of Control Variables

Estimated non-white population in parish, 1911

We first search the full-count release of the 1911 England and Wales census for individuals with birth places in Asia or Africa. In order to filter out births to British subjects on colonial service, we exclude all individuals with a surname that is listed on either `http://www.surnamesdb.com` or `http://www.houseofnames.com`, websites that provide genealogical information on traditional British surnames. Finally, we exclude all individuals with a first name that features in a list of the top 100 boys' and girls' baby names in the 1911 census published online at `http://www.weddingvendors.com/baby-names/popular/1911/`.

Population density in parish, 1931

Population density of the parish containing the base/output area.

Rate of population growth in parish, 1921-1931

Change in parish between 1921 and 1931.

Urban district, 1931

Dummy variable indicating whether the location is in a local government district is an urban (as opposed to rural) district.

Agricultural/professional share, 1931

Share of population employed in agricultural/professional industries, measured at the local government district level.

Unemployment rate, 1931

Unemployment rate, measured at the local government district level.

Distance to coast

Distance to the coast, calculated using R.

Distance to nearest city, 1939

Distance to nearest city with a population of at least 100,000 in 1939, calculated using R.

Distance to nearest large city, 1939

Distance to nearest city with a population of at least 300,000 in 1939, calculated using R.

Distance to nearest urban district, 1931

Distance from base or centroid of output area to the nearest district with urban (as oppposed to rural) status, calculated using R.

Distance to railway station, 1939

Distance to the nearest railway station active in 1939. The Railway Network Inspire dataset provides details on all railway stations in the U.K. in 2016. In order to approximate the railway network in 1939, we supplement this with data from the Wikipedia page 'List of closed railway stations', available at `https://en.wikipedia.org/wiki/List_of_closed_railway_stations_in_Britain`. From this list, we collect geographic coordinates for all railway stations closed after 1939 and append these to the Inspire dataset. Distance to the nearest station from each neighbourhood and base is calculated using R.

Distance to major road, 1946

Calculated based on the basis of a map of the UK road network in 1946 at 1:625,000 scale from the Ordnance Survey Planning Maps series. First, we georeference and digitise the map using ArcGIS. Then we calculate, for each district and base, the distance to the nearest trunk or class I (A) road using R.

Conservative vote share, 1935

Share of voters voting for the Conservative candidate in the 1935 election, measured at the constituency level. Missing if no Conservative candidate stood for election in the constituency.

British Union of Fascists branch

Dummy variable indicating if a location is inside a constituency which contained a branch of the British Union of Fascists, according to internal Labour Party research carried out in June 1934. Missing if the local Labour party branch did not reply to the central party's questionnaire.

Appendix A.4: Mass Observation Directive Responses

"War time events and experiences have changed my outlook very little on this matter. From time to time I have seen coloured soldiers drunk and disorderly in the West End of London. That taught me not to begin to idealise the coloured people but to see that they were human beings, with weaknesses, like ourselves."

"Now for the second part of the question – have wartime events and experiences had any effect on my attitude. The answer is yes. The presence of many more American negroes in this country may make me take less interest them through accepting them as the normal and familiar, they seem a great deal more acceptable to the British public than the American whites."

"When negro American troops first came over here there seemed to be growing up a nasty situation. To read the New Statesmen one would have thought that the presence of negro troops, & the white U.S. forces attitudes towards them, had the makings of a grade A situation; a potential flaw in allied unity. But my experience in East Anglia, Cornwall & the north tells me that friction of any kind has been very rare (or else I just haven't heard about it). What is certain is that U.S. coloured troops have behaved so excellently over here that everybody has good words for them; they like the cheerful grins that come from behind the steering wheels of a convoy of U.S trucks; they like the well-behaved little swagger of negro troops off-duty. During the summer of 1942 there was that army order about keeping aloof from coloured troops to avoid the risk of rows with white U.S. troops. That, I'm glad to say was very unfavourably received by the troops – both non-combatants & royal engineers of the bomb-disposal company in which I was at the time. My girl in the Waafs down in Cornwall looks back with pleasure on the many lifts she'd had into town from the aerodrome by coloured US truck drivers. They would always stop for you, she says, often without being thumbed; would come round to the back of the truck to hoist the girls into the truck & would drop you obligingly at whatever point you wanted to reach. And they were always cheerful & friendly & good-mannered; never 'fresh' like some American soldiers (she said) were."

"I usually get on well with American negroes, who have nicer manners than white Americans."

"I don't think wartime has had any effect on my attitude, expect to make me feel more disgusted than ever with America for her attitude to the negro population. It seems to me quite incomprehensible for a country calling itself democratic and progressive."

"What I hear of USA soldiers & darkies mixing with white women made me furious. After all [negroes only in USA at all on [??] of this of our English fathers. Their place is in Africa."

"I was really shocked (literally) the other day to read of some white girls frequenting a negro military camp. I don't know how they could. Of course I think negroes should have full entry to all hotels etc'."

"Contact with U.S troops in this country has confirmed the fact their attitudes towards black people is revolting and quite undemocratic. It is every bit as justifiable as Jew baiting."

"The N. American negroes are good natured fellows, & several have distinguished themselves well above the average white, but I am horrified at the idea of a British girl marrying a black man. The wars have not changed my opinion on this artefact."

"War time experience of U.S. army negroes has been disheartening. Taking a cross-section, I can't deny I have found them culturally and mentally lower than, say, our Pioneer Corps chaps."

"I do not like the coloured Americans over here & do not think they should have been sent to this country. They are a cause of a great deal of trouble at dance-halls etc. The way girls run after them is a disgrace & in my opinion this little hussies should be spanked."

"White & dark American soldiers were both stationed near a town in which my unit was for some time. The darkies were more charming & less self-assertive. The colour-bar has been made against the wrong colour."

"Wartime has only strengthened my feeling. I hate to think of coloured Americans over here not being treated properly."

"One seldom comes in contact with the Negro African, and although one now sees many American coloured troops, one doesn't come to know them. I suppose I am inclined to think of them as a rather inferior class of American, and I suppose that in the mass they are, despite outstanding figures like Paul Robeson and others'... I do not think that the war has had any effect whatsoever on my outlook on this question, and I have never spoken to a coloured person since its outbreak."

"More recently, knowledge of several American Negro troops has confirmed me in believing that they are in everyway as intelligent and cultured as their white colleagues."

"I do not consider the colour of a persons skin as a matter of any importance at all, and I doubt whether there [are] any innately inferior races. However I was concious of a slight shrinking when a negro soldier came and sat by me in a restaurant before. I have not noticed this before."

"But I have had little actual contact with coloured people. Perhaps I should act like those whom I now criticise. Certainly, it is repugnant to me when I see Leicester

factory girls – ignorant & of low maturity through most of them may be – associating with coloured troops and it gives me rather a shock, or feeling of disgust, when I see a white woman with a coloured child of her own."

"Wartimes experiences or contact with coloured people have been nil, except that I have seen American coloured troops in the district but never spoken to them."

"I have no explicit feeling towards coloured races except perhaps a little against American Negroes. [] The slight feeling of revulsion with Negroes is due to American Negro troops."

"Perhaps my angle on the negroes is partly as it is because of my experience of black troops in this area, though that has merely confirmed previous impressions: it has been found necessary to move these troops constantly, and often at short notice, to avoid outbreaks of violence, while even so the number of rapes and assaults of all kinds has been enormous. []. Lynching becomes explicable."

"My little contact with the American Negroes made me more sympathetic to them. They liked being over here, because they were treated better here by us than by the white Americans in their own country."

"We have very few coloured people in this district. There are one or two American negroes stationed here that I have seen in the streets & a few odd children – negro fathers and white mothers."

"I haven't actually met any coloured people. [] I am furious with the Americans for having such a bad attitude to the negroes."

"The only alteration in my attitude caused by this war is the realisation that the American negro is surprisingly childish. I had been told this before but by coming into contact with the American Negro troops, I have no illusions left as to the negro being equal in mental powers to the white man."

"I have never had anything to do with any negroes. The American negroes who drive the lorries over here look very good tempered."

"I have never come into contact with any of the coloured American troops over here, but they seem well behaved and no worse than the average American soldier."

"Emotionally, the negro American soldier frightens me, I dislike them, I suspect them to whip out a knife and slit my through & definitely do not like to see them with white girls. [] The fact remains I have a considerable emotional prejudice against black men with whole women. The dislike of negroes is due, I think, to a hangover from childhood & would disappear with greater familiarity."

"There are American Negro troops in Belle Vue Manchester. They never seem to be around the city."

"The only wartime event which has had any effect on my attitudes is the tales I hear of US white & black troops, & the reception they have in England. It's a pity to antagonize the US whites by being more courteous to the blacks, but I suppose it's natural to English people who believe the blacks are persecuted in the States."

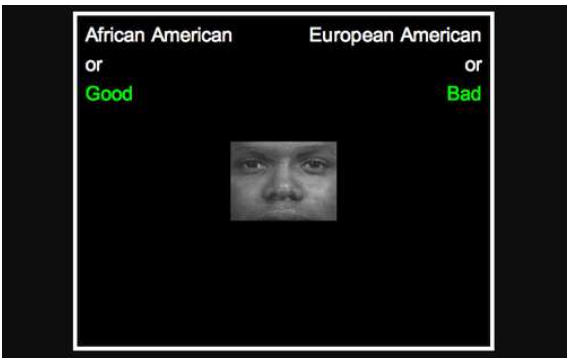
"African negroes I know nothing about & American only from books and recently from seeing their cheerful well conducted troops in the streets of Bristol. Their best can obviously be very good."

"War time events have not had any effect on me, as I have not had anything to do with the black American troops I have had no personal connection with coloured people except two or three Indian students and these were friendly. It was a surprise to find that the first American troops I saw were coloured & not a welcome surprise."

"Yes certainly wartime events have considerably changed my attitudes towards coloured people. There are many coloured soldiers around here. Their 'cause is our cause'."

Appendix A.5: Description of Computerised Implicit Association Test

The race IAT consists of five steps. In each step subjects have to assign a 'stimulus' (a word or a picture of a face) to one of two groups by pressing keys on their keyboard. In step one, subjects sort pictures of black and white faces into the categories 'black' and 'white'. In step two, subject sort words into two categories, 'positive' or 'negative'. The words are all easy to categorise (e.g. "terrible" and "hurt" vs. "joy" and "peace"). In step three, the tasks are combined – for example asking participants to assign black faces and positive words to one category, white faces and negative words to the other. In step four participants again assign faces to categories. Step five is akin to step three, in that participants have to sort both faces and words, but this time the groupings are reversed (the test randomizes whether black faces are shown in step three or step five). So the participant might now have to assign black faces and *negative* words to one category, white faces and positive words to the other. If individuals require more cognitive effort to pair a) black faces and positive words and b) white faces and negative words than a) black faces and negative words and b) white faces and positive words, their response times will vary between blocks three and five. This is measured by the "D" score, the average difference in response times between step three and step five, normalised by the standard deviation of response times in all steps. Higher scores indicate a stronger automatic association of positive words with white faces, with a score over zero indicating implicit racial bias toward whites (Greenwald, McGhee, and Schwartz, 1998).



Screenshot from Project Implicit IAT.

Appendix A.6: Extra Figures and Tables

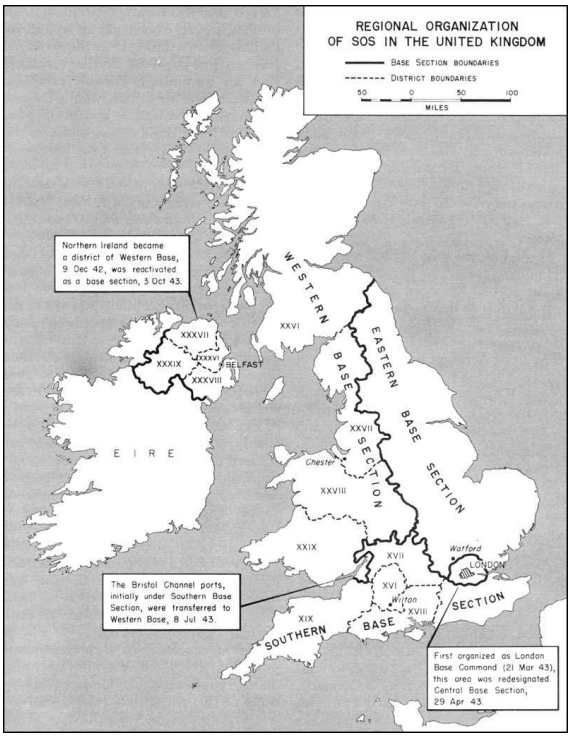


Figure A1. Base Sections in the UK, figure taken from Ruppenthal (1978, p. 85).

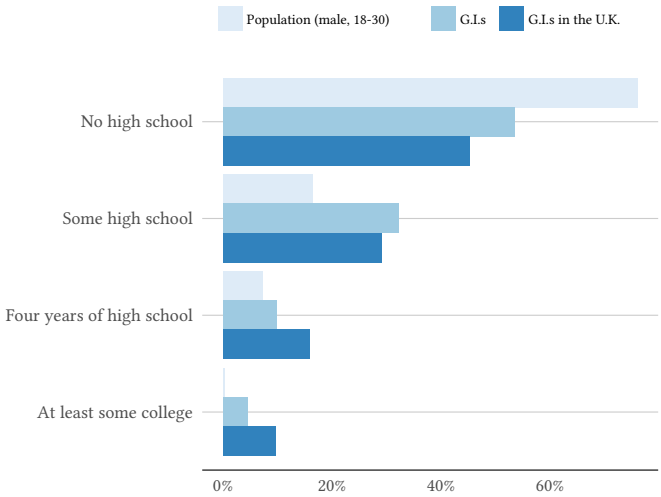


Figure A2. This figure compares the education levels of black males aged 18-30 in the U.S. population as per the 1940 census, G.I.s coded as black in the U.S. Army Enlistment Records to the black G.I.s sampled in the War Department's "Attitudes Towards Army Life" survey (S-92), carried out in the U.K. in November 1943.

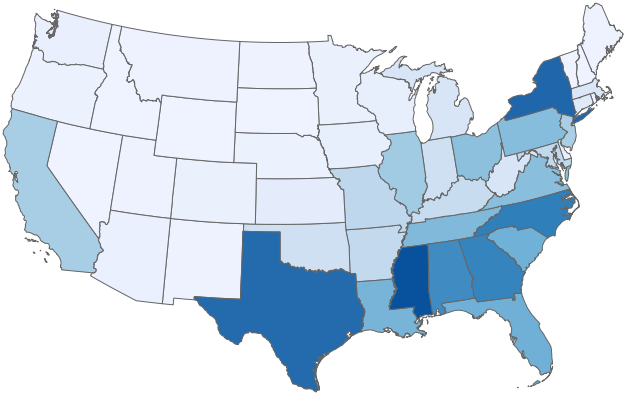


Figure A3. Figure depicts the state of births of enlisted African American G.I.s, according to U.S. Army Enlistment Records. Darker colours indicate more black G.I.s serving from a given state.

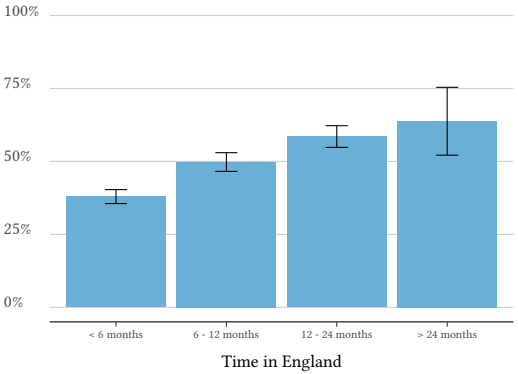
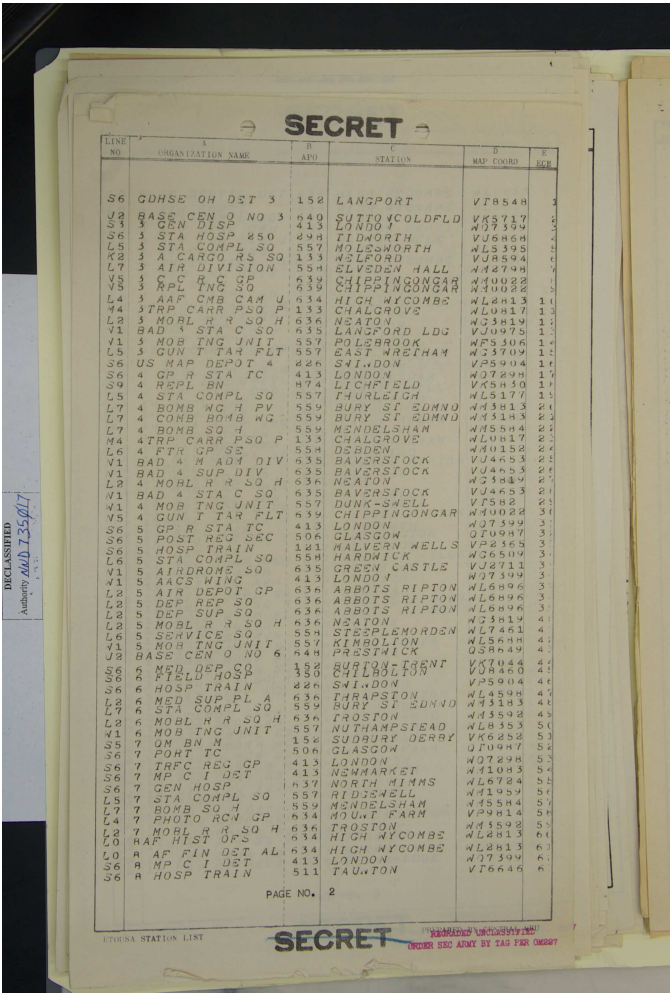


Figure A4. Proportion of soldiers knowing at least some English families or citizens, according to the amount of time spent in Britain. The sample is 3,261 individuals posted in Britain in April 1944. Source: "Attitudes Towards The British" (S-122).



SECRET

LINE NO	A ORGANIZATION NAME	B APO	C STATION	D MAP COORD	E ECH
S6	GDHSE OH DET 3	152	LANGPORT	VT8548	1
J2	BASE CEN D NO 3	640	SUTTON COLDFLD	VK5717	2
S3	3 GEN DISP	413	LONDON	WJ7399	3
S6	3 STA HOSP 250	494	TIDWORTH	VU3464	4
L5	3 STA COMPL SQ	557	MOLESWORTH	WL5395	5
K2	3 A CARGO RS SQ	133	WELFORD	VU8594	6
L7	3 AIR DIVISION	554	ELVEDEN HALL	WM2798	7
V5	3 C C R C GP	639	CHIPPINGONGAR	WM0022	8
V5	3 RPL TNG SQ	639	CHIPPINGONGAR	WM0022	9
L4	3 AAF CMB CAM U	634	HIGH WYCOMBE	WL2813	10
M4	3TRP CARR PSQ P	133	CHALGROVE	WL0817	11
L2	3 MOBL R R SQ H	636	NEATON	WJ3819	12
V1	BAD 1 STA C SQ	635	LANGFORD LDG	VU0975	13
V1	3 MOB TNG UNIT	557	POLEBROOK	WF5306	14
L5	3 GUN T TAR FLT	557	EAST WRETHAM	WJ3709	15
S6	US MAP DEPOT 4	426	SWINDON	VP5904	16
S6	4 GP R STA TC	413	LONDON	WJ7399	17
S9	4 REPL BN	874	LICHFIELD	VK5830	18
L5	4 STA COMPL SQ	557	THURLEIGH	WL5177	19
L7	4 BOMB WG H PV	559	BURY ST EDMND	WM3813	20
L7	4 COMB BOMB WG	559	BURY ST EDMND	WM3183	21
L7	4 BOMB SQ H	559	MENDELSHAM	WM5584	22
M4	4TRP CARR PSQ P	133	CHALGROVE	WL0817	23
L6	4 FTR GP SE	554	DEBDEN	WM0158	24
V1	BAD 4 M AOM DIV	635	BAVERSTOCK	VU4653	25
V1	BAD 4 SUP DIV	635	BAVERSTOCK	VU4653	26
L2	4 MOBL R R SQ H	636	NEATON	WJ3819	27
V1	BAD 4 STA C SQ	635	BAVERSTOCK	VU4653	28
V1	4 MOB TNG UNIT	557	DUNK-WELL	VT5188	29
V5	4 GUN T TAR FLT	639	CHIPPINGONGAR	WM0022	30
S6	5 GP R STA TC	413	LONDON	WJ7399	31
S6	5 POST REG SEC	506	GLASGOW	QJ0947	32
S6	5 HOSP TRAIN	141	MALVERN WELLS	VP2365	33
L6	5 STA COMPL SQ	558	HARDWICK	WG6509	34
V1	5 AIRDROME SQ	635	GREEN CASTLE	VU2711	35
V1	5 AACS WING	413	LONDON	WJ7399	36
L2	5 AIR DEPOT GP	636	ABBOTS RIPTON	WL6896	37
L2	5 DEP REP SQ	636	ABBOTS RIPTON	WL6896	38
L2	5 DEP SUP SQ	636	ABBOTS RIPTON	WL6896	39
L2	5 MOBL R R SQ H	636	NEATON	WJ3819	40
L6	5 SERVICE SQ	554	STEEPLEMORDEN	WL7461	41
V1	5 MOB TNG UNIT	557	KIMBOLTON	WL5688	42
J2	BASE CEN D NO 6	640	PRESTWICK	QS8649	43
S6	6 MED DEP CO	152	BURTON TRENT	VK7044	44
S6	6 FIELD HOSP	450	CHILBOLTON	VU3460	45
S6	6 HOSP TRAIN	426	SWINDON	VP5904	46
L2	6 MED SUP PL A	636	THRAPSTON	WL4598	47
L7	6 STA COMPL SQ	559	BURY ST EDMND	WM3183	48
L2	6 MOBL R R SQ H	636	TROSTON	WM3598	49
V1	6 MOB TNG UNIT	557	NUTHAMPSTEAD	WL8353	50
S5	7 GM BN M	152	SUDBURY DERBY	VK6252	51
S6	7 PORT TC	506	GLASGOW	QJ0947	52
S6	7 TRFC REG GP	413	LONDON	WJ7399	53
S6	7 MP C I DET	413	NEWMARKET	WL1083	54
S6	7 GEN HOSP	637	NORTH MIMMS	WL6724	55
L5	7 STA COMPL SQ	557	RIDGEWELL	WM1959	56
L7	7 BOMB SQ H	559	MENDELSHAM	WM5584	57
L4	7 PHOTO RCN GP	634	MOUNT FARM	VP9814	58
L2	7 MOBL R R SQ H	636	TROSTON	WM3598	59
L0	HAF MTST OPS	634	HIGH WYCOMBE	WL2813	60
L0	A AF FIN DET AL	634	HIGH WYCOMBE	WL2813	61
S6	A MP C I DET	413	LONDON	WJ7399	62
S6	A HOSP TRAIN	511	TAUNTON	VT6646	63

PAGE NO. 2

ETOUSA STATION LIST

SECRET

ORDER SEC ARMY BY TAG PER OM2927

DECLASSIFIED Authority NND 735017

Figure A5. Example station list. Photo taken in the National Archives in Washington D.C.

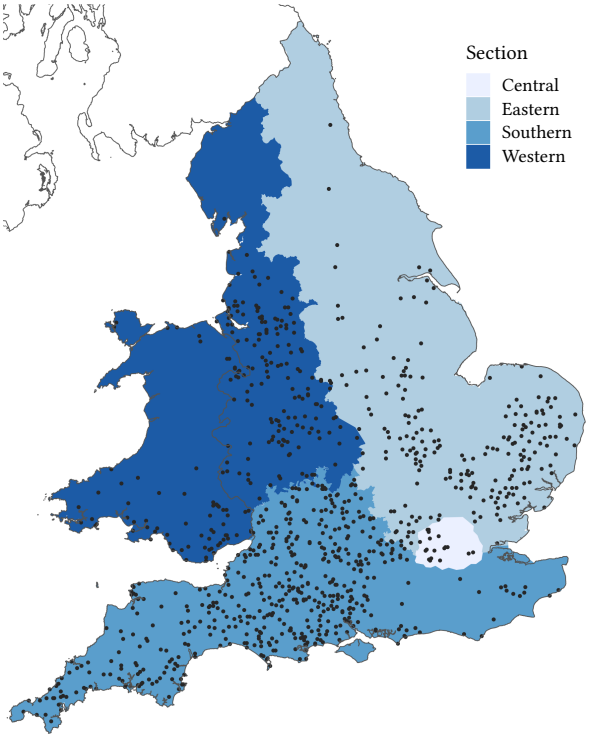


Figure A6. Troop locations across England and Wales. This map shows locations of all locations where, according to our data, U.S. troops were stationed at any point in time during World War II.

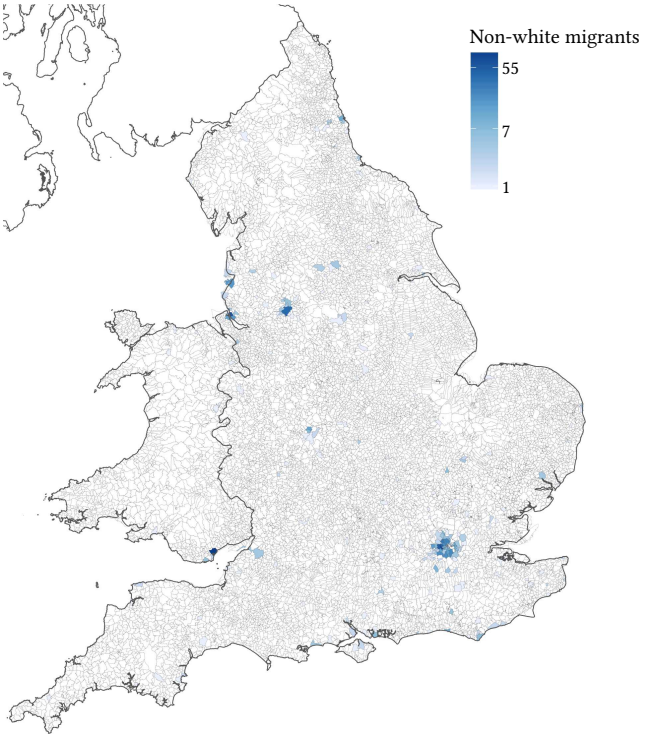


Figure A7. Figure shows the estimated distribution of non-white migrants across parishes in England and Wales, 1911.

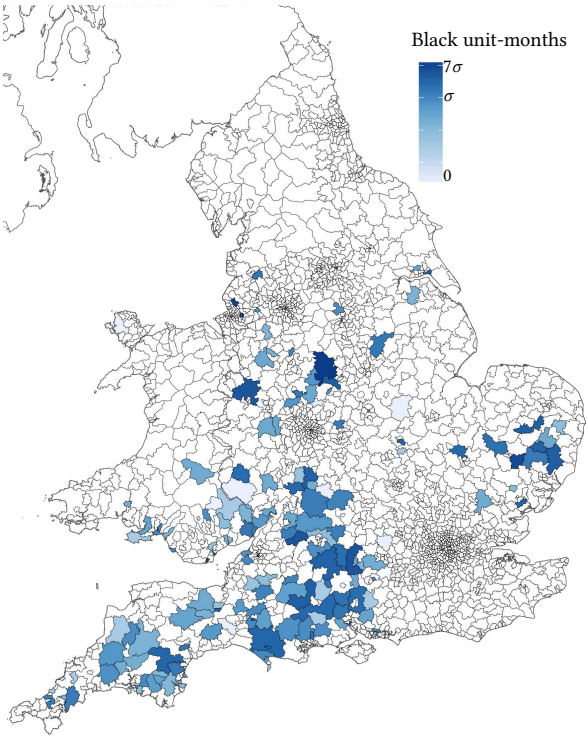


Figure A8. Distribution of $BlackUnitMonths_j$ (standard deviation σ) across postcode districts.

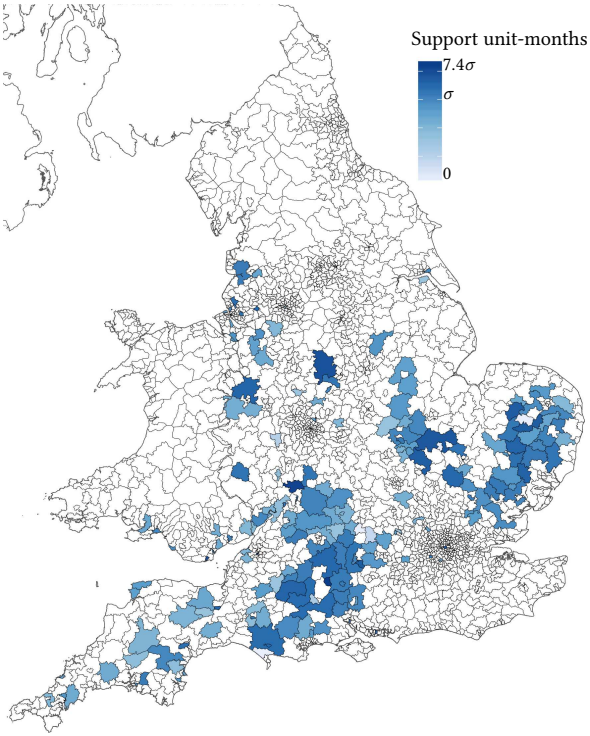


Figure A9. Distribution of $SupportUnitMonths_j$ (standard deviation σ) across postcode districts.

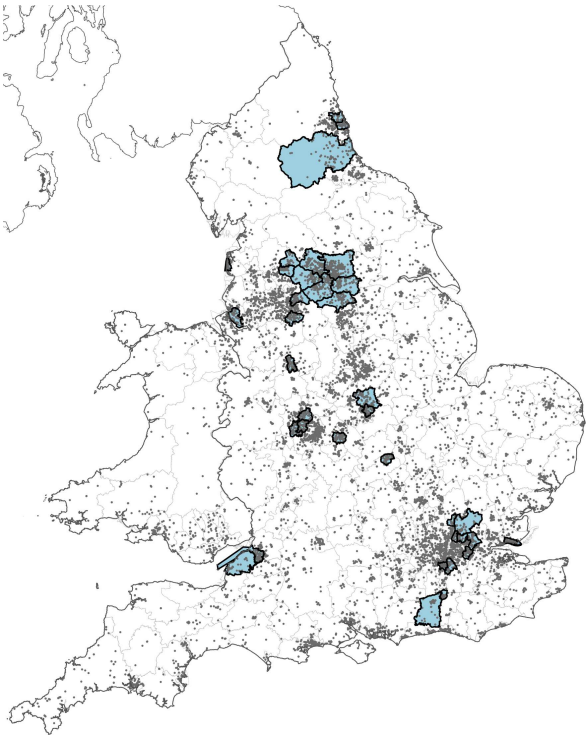


Figure A10. Locations of BNP members and BNP branches.

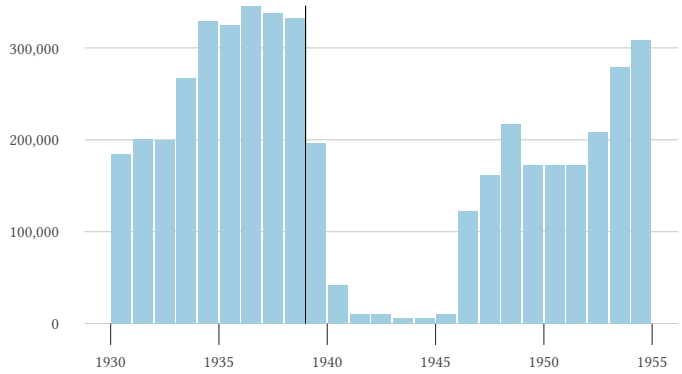


Figure A11. New houses built by year in the U.K. Source: Holmans (2005)

.

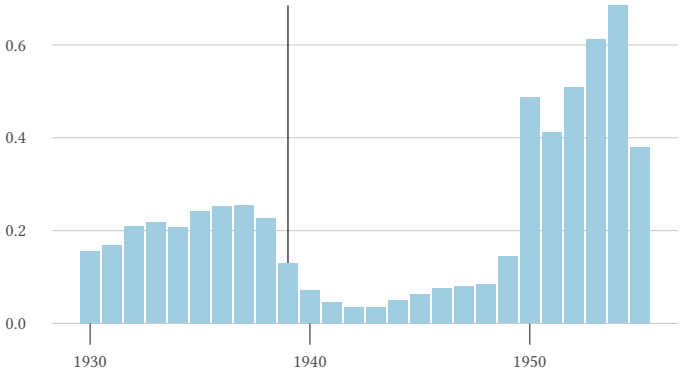


Figure A12. Newspaper housing adverts. We search the British Newspaper Archive (`http://www.britishnewspaperarchive.co.uk/`) for adverts containing the keywords 'semi-detached house', 'detached house' or 'house to let' in each year, and normalise this by the number of adverts containing the keywords 'for sale' or 'to let'.

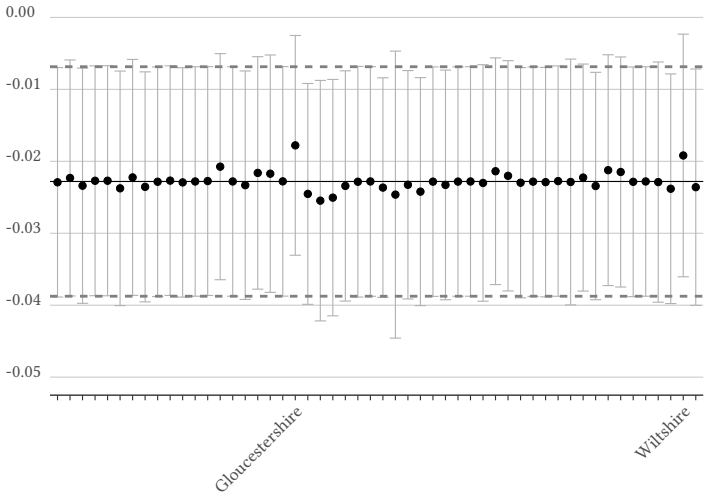


Figure A13. Leave-one-out plot. This figure plots the coefficients (black circles) and 95% confidence intervals from regressions of *BlackUnitMonths* on BNP members per 100,000 whites, at the neighbourhood level. Controls are as per Column 4 of Table 1.2. In each regression, all neighbourhoods inside a given county (shown on the horizontal axis) are excluded from the sample. The estimated coefficient from the baseline specification is shown with a solid black line.

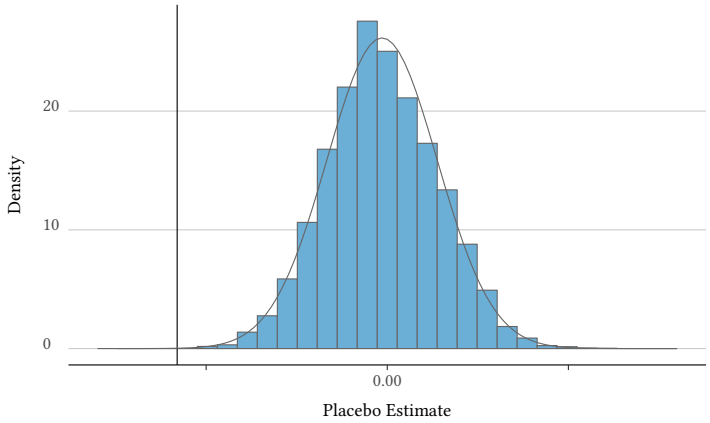


Figure A14. Randomization inference. This figure presents the distribution of coefficients resulting from estimating equation (1) under counter-factual treatments. For each base b, we calculate ($BlackUnitMonths_b$/$SupportUnitMonths_b$), i.e the share of support unit-months due to black units. We then randomly shuffle this value between bases, and create a counterfactual $BlackUnitMonths_b$ for each base by multiplying the shuffled value of ($BlackUnitMonths_b$/$SupportUnitMonths_b$) by the true value of $SupportUnitMonths_b$. We then estimate Equation 1.1, using the counterfactual values of $BlackUnitMonths_b$ and the true values of $SupportUnitMonths_b$ to generate the postcode district-level treatment variables, and estimate Equation 1.1, with controls as per Column 4 of Table 1.2. We repeat the procedure 8,000 times, generating a distribution of counterfactual estimates on $BlackUnitMonths_i$, which is plotted in the figure. The baseline estimate is depicted as a black vertical line, and is larger in magnitude than all but 0.1% of the counterfactual estimates.

Table A1: Changes in opinions caused by American presence in the UK

	Dependent variable:					
	Higher opinion of English			English have higher opinion of Americans		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.28***	0.28***	0.29***	0.21***	0.19***	0.17***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)
Unit controls		✓	✓		✓	✓
Individual controls			✓			✓
Observations	2,560	2,560	2,525	2,503	2,503	2,471

Notes: Columns report results from OLS regressions. The sample is soldiers surveyed in the War Department "Attitudes Toward Army Army Life" survey (S-92), carried out in the U.K. in November 1943. The dependent variable in Columns 1 to 3 is a binary variable indicating if a survey respondent's opinion of the English has become more favourable since being stationed in the United Kingdom. The dependent variable in Columns 4 to 6 is a binary variable indicating if a survey respondent thinks that the English people's opinion of Americans has become higher as a result of having American soldiers in England. Individuals controls are rank, education level, age and state of birth. Unit controls are indicator variables for the branch of the army to which the unit belongs and the station at which the unit is posted. Robust standard errors are reported in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table A2: U.S. Army Station Lists used to construct dataset

Date	Title	Source
21/06/43	MRU United Kingdom Station List	Captain Philip Grinton
14/08/43	MRU United Kingdom Station List	Captain Philip Grinton
30/09/43	MRU United Kingdom Station List	Captain Philip Grinton
14/11/43	MRU United Kingdom Station List	Captain Philip Grinton
31/12/43	MRU United Kingdom Station List	Captain Philip Grinton
21/01/44	MRU United Kingdom Station List	Own digitalisation
21/02/44	MRU United Kingdom Station List	Captain Philip Grinton
31/03/44	MRU United Kingdom Station List	Captain Philip Grinton
30/04/44	MRU United Kingdom Station List	Captain Philip Grinton
31/05/44	MRU United Kingdom Station List	Captain Philip Grinton
30/06/44	MRU United Kingdom Station List	Captain Philip Grinton
14/07/44	MRU United Kingdom Station List	Own digitalisation
31/08/44	MRU United Kingdom Station List	Captain Philip Grinton
31/10/44	MRU United Kingdom Station List	Captain Philip Grinton
25/11/44	MRU United Kingdom and Continental Station List	Captain Philip Grinton
16/12/44	MRU United Kingdom and Continental Station List	Captain Philip Grinton
01/01/45	MRU United Kingdom and Continental Station List	Own digitalisation
02/02/45	MRU United Kingdom and Continental Station List	Own digitalisation
17/04/45	MRU United Kingdom and Continental Station List	Own digitalisation
02/05/45	MRU United Kingdom and Continental Station List	Own digitalisation
04/06/45	MRU United Kingdom and Continental Station List	Own digitalisation
05/07/45	MRU United Kingdom and Continental Station List	Own digitalisation
02/08/45	MRU United Kingdom and Continental Station List	Own digitalisation
04/09/45	MRU United Kingdom and Continental Station List	Own digitalisation
30/09/45	MRU United Kingdom and Continental Station List	Own digitalisation
08/11/45	MRU United Kingdom and Continental Station List	Own digitalisation
11/12/45	MRU United Kingdom and Continental Station List	Own digitalisation
29/12/45	MRU United Kingdom and Continental Station List	Captain Philip Grinton

Table A3: Support units

Unit Role	Black units	Other units	Black share
Quartermaster Truck	317	380	45%
General Service	297	399	43%
Quartermaster Service	239	127	65%
Port Company	176	158	53%
Engineer Battalion Company	149	752	17%
Quartermaster Headquarters	128	150	46%
Quartermaster Other	119	799	13%
Detachment of Patients	98	212	32%
Quartermaster Medical	69	57	55%
General Hospital Complementary	55	114	32%
Ordnance Ammunition Company	48	92	34%
Sterilization	46	46	50%
Signal Construction	45	79	36%
Quartermaster Bakery	41	74	36%
Laundry	38	63	38%
Gas Supply	37	35	51%
Ordnance Base Depot	24	283	8%
Troop Transport	24	0	100%
Fire Fighting	21	169	11%
Quartermaster Railhead	20	38	34%
Engineer Dump Truck Company	17	7	71%
Port Headquarters	17	63	21%
Quartermaster Salvage	14	47	23%
Amphibious Truck Company	12	8	60%
Army Postal	11	328	3%
Ambulance Company	8	49	14%
Chemical Smoke	8	5	62%
Other	216	17,694	0.1%
Total	2,294	22,114	10.3%

Table A4: Effect on BNP membership, treatment at the local goverment district level

	Dependent variable:			
	BNP members per 100,000 whites (std.)			
	(1)	(2)	(3)	(4)
Black unit-months (std.)	-0.016*	-0.024***	-0.019**	-0.016**
	(0.009)	(0.009)	(0.007)	(0.008)
Support unit-months	✓	✓	✓	✓
Grid-cell fixed effects		✓	✓	✓
Economic controls			✓	✓
Geographic controls				✓
Clusters	255	255	254	254
Observations	79,712	79,712	79,496	79,496

Notes: Coefficients from OLS regressions. The unit of observation is the postcode district. Outcome is BNP members per 100,000 white inhabitants. Independent variables are our measure for contact with black troops 'Black unit-months'. Standard errors are clustered at the level of the modern local authority and reported in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table A5: Attitudes towards migration and membership of the BNP, constituency level

	Dependent variable:				
	BNP members per 100,000 whites				
	(1)	(2)	(3)	(4)	(5)
Negative attitude to immigration (std.)	4.99***	5.14***	8.41***	7.68***	7.90***
	(0.46)	(0.47)	(0.74)	(0.60)	(0.85)
Unemployment rate (in %)		1.41***			0.35
		(0.25)			(0.33)
Population density			0.18***		0.039
			(0.027)		(0.035)
Non-white population share (in %)				0.33***	0.27***
				(0.050)	(0.065)
Observations	573	573	573	573	573
R^2	0.11	0.16	0.16	0.20	0.20

Notes: Columns report results from OLS regressions at the constituency level. 'Negative attitude to immigration' is the estimate of average constituency attitudes to migration from Hanretty and Vivyan (2015). The mean of the dependent variable, BNP members per 100,000 whites, is 25. The estimate is based on answers to the question 'Do you think that immigration undermines or enriches Britain's cultural life' in the 2015 British Election Study data, which was answered on a seven point Likert scale. The variable is standardised to have mean 0 and standard deviation one. Population density is measured in persons per hectares. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table A6: Effect on BNP membership, regression at the postcode district level

	Dependent variable:			
	BNP members per 100,000 whites (std.)			
	(1)	(2)	(3)	(4)
Black unit-months (std.)	-0.37***	-0.38***	-0.38***	-0.35***
	(0.10)	(0.10)	(0.10)	(0.1)
Support unit-months	✓	✓	✓	✓
Grid Fixed Effects		✓	✓	✓
1931 population density			✓	✓
Location controls				✓
Conley Standard Error	0.10	0.096	0.096	0.10
Clusters	88	88	88	88
Observations	603	603	603	603

Notes: as per Table 1.2, but on a dataset of postcode-districts generated by aggregating neighbourhoods.

Table A7: Effect on BNP membership, weighted regressions

	Dependent variable:				
	BNP members per 100,000 whites (std.)				
	(1)	(2)	(3)	(4)	(5)
Black unit-months (std.)	-0.021***	-0.023***	-0.022***	-0.023***	-0.029**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.01)
Support unit-months	✓	✓	✓	✓	✓
Grid-cell fixed effects		✓	✓	✓	✓
Economic controls			✓	✓	✓
Location controls				✓	✓
Political controls					✓
Clusters	234	234	234	234	172
Observations	48,732	48,732	48,665	48,665	26,498

Notes: as per Table 1.2, but with observations (neighbourhoods) weighted by the size of their white population.

2

PERSISTENCE AND ACTIVATION OF RIGHT-WING IDEOLOGY

2.1 Introduction

Cultural norms and values—from trust toward strangers, to gender roles, to anti-semitism—often have roots in the distant past and are transmitted across generations.[1] They often persist even when the economic logic that plausibly gave rise to them in the past, in a certain historical and social setting, subsides, or when people migrate into a different country.[2] Undeniably, these norms and values also have a first-order impact on economic outcomes: from female labour force participation, to discrimination, to GDP growth (Tabellini, 2010). And yet, not all historical shocks that shape culture and values manifest themselves up to the present: cultural persistence may be mediated or dampened by intervening factors.[3]

In this paper, we argue that the manifestation of cultural persistence (or lack thereof) is—also—the result of supply and demand factors. Cultural traits may remain dormant:

1. The recent literature in economics on deep roots and persistence of cultural values is large; a very incomplete set of references would include Alesina and Fuchs-Schündeln (2007), Nunn and Wantchekon (2011), Jha (2013), Spolaore and Wacziarg (2013), Alesina, Giuliano, and Nunn (2013), Guiso, Sapienza, and Zingales (2016), Becker et al. (2016), Becker and Pascali (2016).

2. Voigtländer and Voth (2012) show that antisemitism persisted in German cities even after Jews were completely evicted through pogroms; a large literature, using the "epidemiological approach", has shown how second-generation immigrants often mimic the cultural traits and attitudes of their ancestors' country of origin (Giuliano, 2007; Fernández and Fogli, 2009; Alesina and Giuliano, 2010; Algan and Cahuc, 2010; Luttmer and Singhal, 2011).

3. Voigtländer and Voth (2012) show that in some German cities, e.g. those with a tradition of commerce, the transmission of antisemitism is lower. Giuliano and Nunn (2017) provide a broader framework to understand cultural transmission.

present in the population, but not manifest because their expression is too costly. Only as external circumstances change, these traits may resurface again and result in actual social (political, economic) outcomes. Specifically, we look at the political domain, where we investigate the persistence of right-wing ideology in Germany. In recent years, the "Alternative for Germany" (*Alternative für Deutschland*, henceforth AfD) has offered a political platform on the far right: conservative, nationalistic, and at times outright xenophobic. We show that municipalities that, in the 1920s and 30s, expressed strong support for the Nazi party (the NSDAP) now have a stronger vote base for the AfD. In our baseline specification, a one standard deviation increase in Nazi support during the Weimar era is associated with 0.16 standard deviations more support for the AfD in the 2016/2017 state elections. This result is robust to controlling for state fixed effects, and for a host of plausible economic and social determinants of electoral outcomes.

Crucially, the specific setting of Germany allows us to observe a case in which a change in the supply of political platforms is key in making a long-run persistence of ideological traits reemerge. After the catastrophic experience of Nazism and World War II, and the subsequent denazification process, the postwar legal setting severely constrained the expression of right-wing ideology, and almost completely impeded the creation of parties on the extreme right fringe. The "Alternative for Germany" bypassed these constraints: it was founded in 2013 as a platform to promote fiscally conservative principles and oppose the Greek bailout. Two years later, in 2015, the initial group of founders (many of them economists) was ousted and the party veered strongly to the right, focusing on immigration and nationalism (a near-taboo topic in Germany) as main themes. Consistent with our hypothesis, we find that there is only a small or no correlation between the AfD's electoral fortunes and Nazi support in 2013/14, when the AfD espoused merely economic conservatism, whereas a correlation emerges once the political platform is changed after 2015.

Using quantitative language analysis, we document how the AfD markedly changed its rhetoric after its sudden change of leadership in 2015. Instead of focusing on Greece and the Euro crisis, the AfD now emphasised topics such as nationalism and (the perceived threat of) Islam. Notably, only the rhetoric changed—the name and the logo of the party remained identical. This allowed the party to avoid the intense legal scrutiny, and public stigma, that newly-founded right-wing parties are subject to in Germany. The change in rhetoric after 2015 sufficed as a "dog whistle" to make the historical correlation with Nazi support emerge.

The historical correlation we observe is positive, significant, and large. We compare votes cast for the AfD in the federal election 2013 to the local elections that occurred in six German states in 2016/17; in our preferred specification, a one standard deviation increase in historical support for the Nazi party is associated with a 0.15 standard

deviations larger change in votes towards the AfD from 2013 to 2016/17. These results are also robust to a variety of other control variables relating to demographic, religious, and economic characteristics of the municipalities in the analysis; we consider both historical (relating to the 1920s and 30s) and contemporary control variables, such as levels and changes of unemployment.

A major political turning point in 2015 was the sudden, massive inflow of Syrian refugees in September and the following months. This occurred soon after the change in leadership at the AfD's helm, and undeniably influenced the political discourse. Still, we do not think that this political event can explain away our findings. First, the AfD's nationalist turn occurred months *before* the inflow of Syrian refugees (actually, at the peak of the Greek bailout crisis), not as a consequence of it. Second, our empirical strategy rests on a comparison of municipalities and their *change* in voting behaviour between 2013 (when the AfD was merely fiscally conservative, and the refugee crisis had not occurred yet) and 2016/17. To the extent that the Syrian refugee crisis has increased xenophobia and anti-immigration sentiment in Germany overall, this level effect should be captured by considering first differences in voting outcomes. Last, we explicitly control for the presence of refugees in a municipality, and observe that the actual presence of asylum seekers does not sway voters significantly.

Finally, we consider plausible patterns of voter migration. Our analysis suggests that areas which traditionally supported the Nazi party witnessed an increase in voter turnout as the AfD turned more radical; voter mobilisation may thus explain parts of the AfD's electoral success. Along the intensive margin, our analysis reveals that municipalities with high Nazi vote shares in the Weimar era moved away from parties on the far left and far right between 2013 and 2016/17, casting more votes for the AfD instead.[4]

Our analysis speaks to several research agendas in economics and political science. First, and obviously, we contribute to the literature cited above on the long-term persistence of cultural traits and attitudes. As in, e.g., the papers by Voigtländer and Voth (2012), Guiso, Sapienza, and Zingales (2016), Becker et al. (2016), we show that cultural traits—in our specific case, political attitudes—have deep origins that may trace back to the distant past, and be transmitted across generations.[5]

The AfD's electoral successes show, however, that the historical persistence of political attitudes is not always visible, and may need to be "activated" by changes in the institutional setting or the political marketplace. This activation of historical memories

4. In future work, we plan to complement our analysis of electoral results with a parallel analysis of political attitudes as expressed in opinion surveys.

5. Complementary to our context, an important literature in political science has studied the intergenerational transmission of political preferences, see e.g. Beck and Jennings (1991) and Jennings, Stoker, and Bowers (2009). Specifically for the context of right-wing ideology, see Avdeenko and Siedler (2017).

has also been evidenced by two recent papers. In Fisman, Hamao, and Wang (2014), anti-Japanese hatred is selectively stoked by Chinese leaders for domestic policy purposes; the authors show that this reactivation of hatred has an impact on stock market prices of Japanese firms with exposure to China. Fouka and Voth (2016) show how sales of German cars declined, as the recent Greek debt crisis worsened, in Greek localities that witnessed massacres perpetrated by German forces, again during World War II.

In these papers, incidental changes in the political background have economic consequences. Another literature has focused on the endogenous choice of politicians to selectively activate feelings in the electorate: Glaeser (2005) discusses how hatred, e.g. antisemitism, can be strategically used by politicians to mobilize voters. The recent work by Ochsner and Roesel (2017) studies a context close to ours—the populist right-wing FPÖ party in Austria—showing that it is successful in unearthing a resentment against Turkish immigrants that dates back to the Turkish sieges of Vienna in the 16th and 17th century.

Second, our work is a contribution to understanding the determinants of (radical) right-wing voting. In recent years, discontent about the social and economic situation has increasingly expressed itself in votes for candidates and parties on the populist (far) right, from Viktor Orbán in Hungary to the FPÖ in Austria, from Marine Le Pen in France to the UKIP in Britain, and to Donald Trump. Economic insecurity, spurred by increasing globalisation and the demise of traditional manufacturing, may explain part of this political shift.[6] Increasing levels of immigrant population in (Western European) countries may also explain some part of the right's electoral successes.[7] However, closer to the setting studied in our analysis, Steinmayr (2017) finds that the short-term effects of direct exposure to Syrian refugees are more favourable to parties supporting immigration, rather than to xenophobic movements. Finally, Inglehart and Norris (2016) argue that the recent rise of populism can best be understood as a reactionary response to a cultural change that is perceived as too fast and unsettling by some sectors of the population.

This latter research is part of a longer tradition of political scientists trying to understand the emergence of far right parties, especially in a comparative dimension.[8] To our knowledge, we are among the first to bring two new factors, and the interaction thereof, to the explanation of the electoral successes of right-wing parties. On the one hand, we shed light on the role of long-standing, deeply ingrained political beliefs—

6. This important dynamic has recently been studied for the US (Autor et al., 2016), for Germany (Dippel, Gold, and Heblich, 2016), and for France (Malgouyres, 2017).

7. Halla, Wagner, and Zweimüller (2016) and Dustmann, Vasiljeva, and Piil Damm (2016).

8. See, e.g., the recent review by Golder (2016), as well as the earlier works by Norris (2005), Mudde (2007), and Arzheimer (2008).

this is especially salient in Germany, a country that experienced a most destructive instance of fascism.[9] On the other hand, we emphasise the importance of political structures, and the incentives/constraints they set for the formation of political parties, in facilitating the expression of right-wing ideology.[10]

The paper proceeds as follows. In Section 2.2, we provide an introduction to the political context in Germany, and some information about the history of the AfD. In Section 2.3, we describe the data used. In Section 2.4, we use language analysis to quantify the change in the AfD's rhetoric after March 2015. In Section 2.5, we present the empirical analysis linking historical support for the NSDAP with the AfD's electoral results. Finally, in Section 2.6 we conclude.

2.2 Historical Context

2.2.1 The Political Landscape in Germany

After the collapse of the Nazi regime and Germany's defeat in World War II, the reconstruction of the political party system in West Germany (the Federal Republic of Germany, founded in 1949) faced two major challenges. First, parties tried to rebuild a system that would supersede the structural weaknesses of parties during the Weimar era—a weakness that arguably resulted in the end of democracy and the Nazis' takeover of power. Second, parties struggled to integrate large swaths of the population who were actively involved in the Nazi dictatorship, among these an estimated 8.5 million former card-carrying NSDAP party members, into the new democratic system.

On the right side of the political spectrum, the main actor was the Christian Democratic Union (CDU). Founded by several members of the Nazi resistance, it built on the previous experience of the Catholic "Zentrum" party, but explicitly tried to appeal also to Protestant voters, who before the war largely supported nationalist/conservative parties. The CDU (and its Bavarian sister party, the CSU) succeeded in the endeavour of becoming the main conservative party in Germany, channelling nationalists, economic liberals, and social conservatives alike into one party strongly supporting democratic values in the new Federal Republic of Germany. Smaller parties on the right appealing to specific constituencies, such as the BHE (League of Expellees),

9. Despite the availability of high-quality electoral data from the Weimar era, only few researchers have tried to correlate post-war political outcomes in the Federal Republic of Germany with early Nazi support: Liepelt (1967) showed that in 1966 there was a strong correlation between electoral successes of the NPD (a neo-Nazi party) and the NSDAP in 1932. Cf. also the early contributions by Kaltefleiter (1966), Kühnl, Rilling, and Sager (1969), Sahner (1972), and Winkler (1994).

10. Closer to our argument, Arzheimer and Carter (2006) emphasise the importance of structural factors that shape the political opportunity structure and thus the comparative emergence of right-wing movements, as well as the interaction of economic factors and political structures (Arzheimer, 2009).

targeting the expellees losing their ancestral homelands after WWII, and the DP (German Party), appealing especially to war veterans and northern German conservatives, quickly disappeared and were not represented in the federal parliament after 1957.

Other political parties emerging to the right of the CDU/CSU in later years were similarly unsuccessful, scoring at best very temporary successes. The NPD (National Democratic Party) was founded in 1964 and enjoyed some temporary popularity in the late 1960, winning some seats in state assemblies. However, it never managed to break through the 5% threshold of votes required to gain representation in the *Bundestag* (the federal parliament). In the following decades, it drifted further to the right and lost electoral support, only to regain popularity after Reunification in the states of former East Germany. There, it entered two state assemblies (in Mecklenburg and in Saxony) twice in the late 2000s, securing up to 9.2% of the vote. At the federal level, the NPD remained insignificant, never scoring more than 1.6% of the vote after 1990.

Among other attempts of party to break through on the right side of the CDU/CSU, the *Republikaner* (Republicans) were notable for their successes in the late 1980s and early 1990s. They entered two state assemblies, sent MPs to the European Parliament in 1989, but they, too, repeatedly failed to pass the 5% threshold to enter the *Bundestag* and drifted into irrelevance. The ability of the CDU/CSU to squeeze out all margins on the right end of the political spectrum, all the while remaining solidly grounded in democratic and liberal principles, is well summarised by the long-time leader of the CSU, Franz Josef Strauss, who quipped in 1986 that there "shall not be a democratically legitimate party to the right of the CSU."

Indeed, no explicitly nationalistic/right-wing party was ever able to enjoy substantial and continued electoral success in the history of the Federal Republic of Germany. However, this was not only due to the CDU/CSU's ability to occupy the space on the political right: a major factor was also a provision in the Basic Law (the constitution of the Federal Republic of Germany) that enabled the Constitutional Court to disband extremist parties on the left and the right. Article 21.2 of the Basic Law states that *"[p]arties that, by reason of their aims or the behaviour of their adherents, seek to undermine or abolish the free democratic basic order or to endanger the existence of the Federal Republic of Germany shall be unconstitutional."* This article was invoked twice with success: in 1952, the Constitutional Court outlawed the SRP (*Sozialistische Reichspartei*, Socialist Reich Party), a party that had an openly neo-Fascist agenda and recruited former Nazi functionaries, and in 1956 the communist party (KPD).

This provision in the Basic Law was successful in disciplining the extremeness of right-wing political platform even when it did not result in an explicit party ban—the mere threat of disbandment sufficed. The case of the NPD was twice brought to the Constitutional Court, once in the early 2000s, when it was dismissed on formal grounds, and once in 2016-17, when the court ruled that, while the party's ideology is

unconstitutional, its support is too small to undermine the democratic order and thus to justify its ban.

2.2.2 The "Alternative for Germany" (AfD)

In September 2012 three individuals—Bernd Lucke (an economics professor from Hamburg), a former CDU politician, and a journalist—signed an appeal to oppose the current policies pursued by the German government to fight the Euro crisis. The manifesto called for the foundation of a party to be called "Electoral Alternative 2013". Notably, this initial manifesto was only concerned with the Euro crisis, the potential bailout of Southern European states, and the fear that the federal government might cede more powers to Brussels (also suggesting the use of referenda)—the manifesto explicitly ruled out the "alternative" taking a stance on other policy concerns.

In the following months, the "Electoral Alternative" morphed into a fully-fledged party, attracting, in particular, a large number of disappointed former CDU and FDP members, as well as several economics professors. In the 2013 federal election, the "Alternative for Germany" (*Alternative für Deutschland*, or AfD), as it was now named, won 4.7% of the votes, only narrowly missing the 5% threshold to enter the *Bundestag*.

Following the federal election, the AfD gained further strength, obtaining 7.1% of the votes in the European Parliament election of May 2014. This expansion, however, also meant that new party members were not only concerned about Euro crisis policies; in fact, the AfD increasingly attracted conservatives of all sorts, who did not feel represented by the centrist course of chancellor Angela Merkel. The tensions between the initial group of party members—economics professors and fiscal conservatives—and the newer, national-conservative, anti-immigration members became virulent in the spring of 2015, when two leading party functionaries from Eastern Germany published the "Erfurt Resolution."

In this document, they called for a policy of opposition to the "social experiments of the past decades (gender mainstreaming, multiculturalism) […]" and encouraged the party leadership to embrace the xenophobic, anti-immigrant PEGIDA ("Patriotic Europeans Against the Islamisation of the West") movement. At the following party congress in Essen, in July 2015, Frauke Petry, representing the conservative, anti-immigrant wing was unexpectedly elected party leader with 60% of the vote, with only 38% of the votes going to the party founder Bernd Lucke. The congress in Essen sanctioned the takeover of the party by its conservative faction; the fiscal conservatives rallying around Bernd Lucke left the party and founded another movement (with little electoral fortune so far).

With its new, anti-immigrant conservative rhetoric, the AfD enjoyed considerable successes in the state elections held in 2016 in the states of Baden-Württemberg (15.1%),

Rhineland-Palatinate (12.6%), Saxony-Anhalt (24.3%), Mecklenburg-Anterior Pomerania (20.8%), and Berlin (14.2%). The sudden, large expansion of the party, and the election of inexperienced and insufficiently vetted candidates to local assemblies also meant that the AfD was involved in several political scandals, especially with regard to extremist political views (antisemitism, holocaust denial). The party leadership also moved further to the right, ousting Frauke Petry in the 2017 national congress and replacing her with even more conservative members.

Perhaps as a consequence of this further radicalization, the party's performance in local elections in 2017 was less impressive; still, it gained representation in all state assemblies that were up for election in that year (Saarland, 6.2%, Schleswig-Holstein, 5.9%, North Rhine-Westphalia, 7.4%). As we write, the AfD is preparing for a federal election (scheduled in September 2017) with a programmeme that espouses staunchly conservative values (law and order, traditional family values), but also less established views (climate change denial, scepticism of mainstream media) and bordering outright xenophobia (calling for a stop to immigration, especially of asylum seekers and family reunifications, limiting access of immigrants to social security, and demanding German values rather than a multicultural society).[11] Figure 2.1 reports a timeline of the major events.

2.3 Data Description

Our body of data used in this research consists of three parts: *(i)* electoral data; *(ii)* data documenting the relative shift in political platforms, as reflected by political language; and *(iii)* other control variables.

2.3.1 Electoral Data

Our electoral data are drawn from the official website of the Federal Returning Officer (*Bundeswahlleiter*) for the federal election to the *Bundestag* in September 2013, or from the websites of the respective state returning officers (*Landeswahlleiter*) for the elections to state parliaments (*Landtage*) in 2016 and 2017. The data are provided at the municipality (*Gemeinde*) level.

Data for the federal (*Bundestag*) elections prior to 2013 are obtained from DESTATIS, the German federal statistical office. We purchased the municipality-level tabulations of all elections from 1989 until 2009 (data for the elections in 1980, 1983, and 1987 refer to West Germany only).

11. See "Programm für die Wahl zum Deutschen Bundestag am 24. September 2017", `https://www.afd.de/wp-content/uploads/sites/111/2017/06/2017-06-01_AfD-Bundestagswahlprogrammemem_Onlinefassung.pdf`, last accessed 14 August 2017. Hensel et al. (2017) provide a comprehensive overview of the AfD's history and current goals.

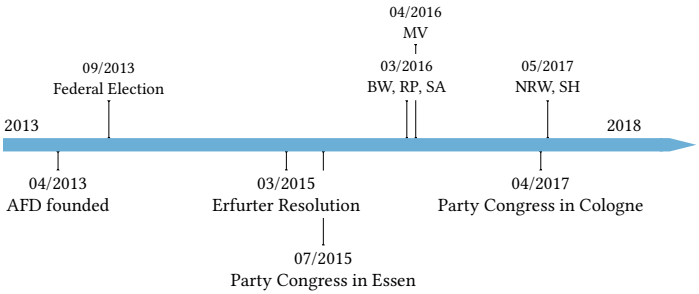


Figure 2.1. Dates of major elections (federal elections and elections to state parliaments) included in our analysis are reported above the timeline. These comprise elections to the legislatures of Baden-Württemberg (BW), Rhineland-Palatinate (RP), Saxony-Anhalt (SA), Mecklenburg-Anterior Pomerania (MV), North Rhine-Westphalia (NW), and Schleswig-Holstein (SH).

For the electoral results of right-wing parties during the Weimar Republic, we make use of the pathbreaking work of Jürgen Falter and Dirk Hänisch (Falter and Hänisch, 1990), who digitised the votes for the *Reichstag* elections from 1920 until 1933 as published in the series *Statistik des Deutschen Reiches*. In all years, except for the two elections of 1932 (July and November), electoral results were published at the level of counties as a whole (*Kreis* or *Stadtkreis*), and then separately for all municipalities above 2,000 inhabitants contained in a county.[12] From this disaggregation, we can easily reconstruct the aggregate votes for all municipalities contained in a county, but below the 2,000 inhabitants threshold (the "remainder of the county").

We match present-day electoral outcomes to the Weimar era party support through a geocoding algorithm, in two steps: in the first step, we geocode the Weimar-era electoral entities (counties and municipalities) listed in the Falter and Hänisch (1990) dataset, using a combination of historical county shape files,[13] current geodata from OpenStreetMaps, and a variety of other online sources. In the second step, we match modern electoral geographies to these geocoded entities. Based on the geographic location, a current municipality is either matched to a city-county (*Stadtkreis*) of the Weimar era, or to one of the municipalities whose electoral data is known because it

12. For the elections of 1932, no data at a level of disaggregation below the county were published. After 1933, the new regime unfortunately had priorities other than publishing past electoral results. We therefore cannot use the 1932 electoral results in our analysis.

13. Provided through the Census Mosaic project, `http://www.censusmosaic.org`.

had more than 2,000 inhabitants. We call these municipalities "exact matches". The remaining municipalities are then assigned, based on their location, to the entity "remainder of the county", i.e. to the aggregate electoral results in a historical county, *outside* the municipalities with more than 2,000 inhabitants. Typically, for any Weimar-era observation relating to the "remainder of the county", there will be several present-day municipalities matched. We account for this by clustering our regression analysis at the level of observation in the Weimar era (*Stadtkreis*, municipality above 2,000 inhabitants, or "remainder of the county").[14]

2.3.2 Text Data

We analyse the language used by the AfD and other major German parties by considering a variety of sources. Besides the AfD, we consider the following parties: CDU/CSU[15] (Christian democratic, moderately conservative), SPD (social democratic, moderately left-wing), Grüne (green party), FDP (free democrats, economic/socially liberal), and the NPD (nationalistic, starkly right-wing, only represented in a few state legislatures).

First, we look at all party manifestos, official documents setting out the parties political platforms in advance of major elections (federal, state, and European Parliament elections), published from 2013 onwards. These manifestos are usually published a few months ahead of the election, and contain variously detailed statements of political objectives and policy proposals. We obtained the full text (as PDF) of 74 manifestos from the respective party websites; the median manifesto is 56 pages long and encompasses approximately 19,500 words.

Second, we consider the content of major political speeches held at party congresses, at national-level party meetings (e.g., the traditional Epiphany meeting of the FDP on 6 January), or so-called "Ash Wednesday" speeches[16] by major political leaders (usually the party secretaries or the main candidates), from 2013 until today. If the speeches are not available in a transcribed version, we resort to online videos of these speeches and transcribe them with speech recognition software or manually. Our final dataset contains 112 speeches; the median length of a speech is 27 minutes.

Third, we analyse tweets posted from the official Twitter accounts of those six major parties (we restrict ourselves to the main/national account of the party, not of its regional branches and candidates). We scrape all tweets from April 2008 (when the

14. Supplementary Appendix B.1 describes this algorithm in detail.

15. For speeches and party manifestos, we consider the CDU and the CSU as one party (among other reasons, because of the low number of observations). For tweets and Facebook posts, we look at the CDU and the CSU accounts separately.

16. On Ash Wednesday, all major political parties in Germany hold speeches, often in beer halls, which are typically more polemical and more directly targeted against opponents.

first party, the CDU, opened a Twitter account) until the end of June 2017, obtaining a total of 66,422 tweets (the most prolific party is the NPD, with 18,057 tweets, followed by the SPD, with 10,580 tweets; the AFD posted 4,119 tweets).

Finally, we also scrape posts from the official Facebook pages of the major parties (again restricting ourselves to the federal-level party organisation, not to its local branches). We obtain 36,089 posts from November 2008 until May 2017; 12,794 of these posts pertain to the NPD page, 2,881 to the AFD.

2.3.3 Other Variables

We complement our analysis of electoral results with a range of both historical and contemporary control variables—variables that may be potentially omitted factors in our regression setups. For the Weimar era, we use the same dataset by Falter and Hänisch (1990), which also contains statistics on, among others, population, unemployment, employment structure, and religious composition in 1925 and 1933. Population and religion data are available at the municipal level (municipalities above 2,000 inhabitants); all other statistics are only available at the county level. We match those statistics to contemporary voting outcomes using the same algorithm as for electoral data.

We also include a variety of contemporary control variables in our electoral data regressions. These comprise the unemployment rate, total population, male population share, and area of the municipality (all measured in 2013) in our regressions, as well as a full set of dummies characterizing the degree of urbanisation of a municipality.[17] These data are obtained from DESTATIS.

Note, however, that our preferred specification focuses on the *change* in electoral support for the AfD between 2013 and the following state election (in 2016 or 2017). Therefore, one should be particularly interested in the potentially confounding effect of other changes occurring in the same time frame. For example, we obtained data on the change in unemployment, at the county level, between 2013 and 2016 (from DESTATIS). The most salient political event happening in this time frame is the "(Syrian) refugee crisis", which peaked in the fall of 2015 after Germany's decision to suspend the Dublin agreement and not to deport asylum seekers back to the first EU member state they entered. While most asylum seekers enter Germany through the German-Austrian border in the south-east of the country, they are supposed to be reallocated to the single federal states, and then again to counties, according to a quota system which takes into account population and GDP. Within counties, asylum seekers are

17. Following EUROSTAT guidelines, DESTATIS classifies municipalities according to its urbanisation density as follows: "densely populated" if at least 50% of the population lives in high-density clusters, "thinly populated" if more than 50% of the population lives in rural grid cells, and "intermediate density" (all other municipalities).

further assigned to municipalities according to a variety of criteria. From the Federal Employment Agency (*Bundesagentur für Arbeit*), we obtain the number of asylum seekers in each municipality, as of December 31, 2016.[18] Given the low number of refugees in 2013, we consider the *level* of asylum seeker in December 2016 as a close approximation to the change relative to 2013.

2.4 Evidence on Semantic Change

We view the turn of the AfD from a monothematic, anti-Euro and anti-Greek bailout party to a more traditional xenophobic, anti-immigrant right-wing party as a suitable policy experiment in which an existing party changes its placement on the political spectrum, without changing the name, logo, or most of the party structures.[19] Clearly, this change was also perceived by the voters. In the surveys conducted for the German Longitudinal Election Study (GLES), potential voters were asked to place parties on an 11-point left-right scale.[20] As shown in Figure 2.2, in 2013 voters were not sure where to place the AfD on a left-right spectrum; the modal answer (excluding "I don't know", which is by far the preferred response) is the score of 6, right in the middle of the spectrum, and the median is 7, just to the right of the centre. Over the course of the following years, the public perception of the party shifted radically, especially after 2015. In 2016, only few voters cannot place the AfD on a left-right spectrum, and most place the party to the far right (the rightmost answer, 11, is also the modal answer).

We back up this survey evidence with text data. To do this, we analyse the language used by the AfD and by other major German parties on different channels: in party manifestos, in major speeches, in Facebook posts (on the official Facebook pages), and in tweets (on the official Twitter accounts of the respective parties). Figure 2.3 gives a first quantitative impression of the nationalistic turn imparted on the AfD starting in mid-2015.

We classify Facebook posts (looking at trimmed word stems) depending on whether they contain a word that is related to the Euro, to Greece (likely in the context of the bailout talks), to Islam/Muslims, or to Germany/the nation. Up until 2015, about 20% of posts refer, on average, to the Euro, and approximately the same amount refer to Germany/the nation. There is, however, already a slight downward trend in references

18. To be precise, the data from the Federal Employment Agency refer to *Erwerbsfähige Leistungsberechtigte im Kontext von Fluchtmigration*, i.e. potential transfer recipients, able to work, in the context of escape migration. This includes, roughly, all asylum applicants who are above age 15, not disabled, excluding family members who joint first emigrants at a later stage. The exact number of asylum applicants cannot be obtained at the municipality level.

19. Appendix Figures B1 and B2 show, anecdotally, how this change was reflected in party billboards.

20. We use component 8 of the GLES (Long-term online tracking), studies ZA5720, ZA5726, ZA5728, ZA5732. All studies are available through the GESIS website (`www.gesis.org`).

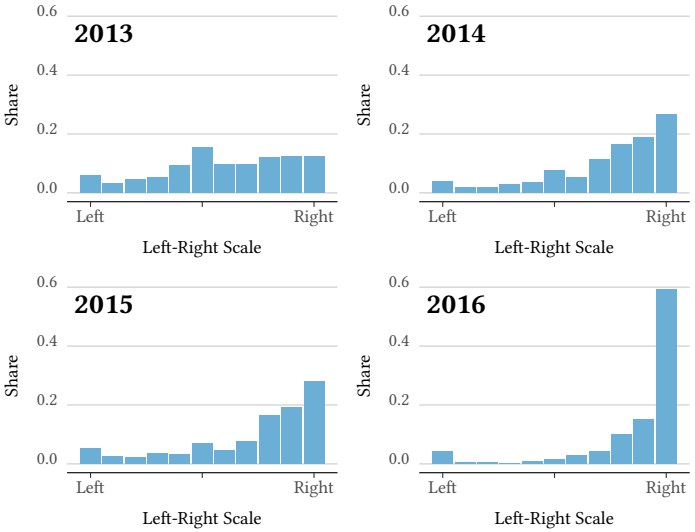


Figure 2.2. The four graphs report the perceived location of the AfD on the left-right political spectrum, as derived from answers to the German Longitudinal Election Study (GLES).

to the Euro before 2015, which suggests that, as the base expanded, the party's outlook widened beyond its initial narrow focus on economic topics. 2015 witnesses two major changes. First, as the Greek crisis approached a new zenith (the infamous "bailout" referendum was held on July 5), Greece and the Euro reach a short-lived peak in frequencies. At the same time, after the party congress in Essen, the AfD turns rightward: posts referring to Germany or the nation steadily increase in frequency, and so do posts referring to Islam or the Muslim world. Note that the latter change only occurs in mid-2016, well after the peak of the refugee crisis in September 2015.

However, these suggestive trends may also be misleading, and merely capture an overall change in topics relevant for German politics. It is plausible that other parties in Germany, in the context of the dramatic political and economic crises of the past years, have readjusted their rhetoric and the focus of their policy proposals. For this purpose, in Table 2.1 we look at the overall text body that we collected in manifestos, speeches, tweets and Facebook posts, for seven major parties in Germany: the AfD, as well as the CDU/CSU, SPD, Greens, FDP, Linke, and (as a benchmark of a more

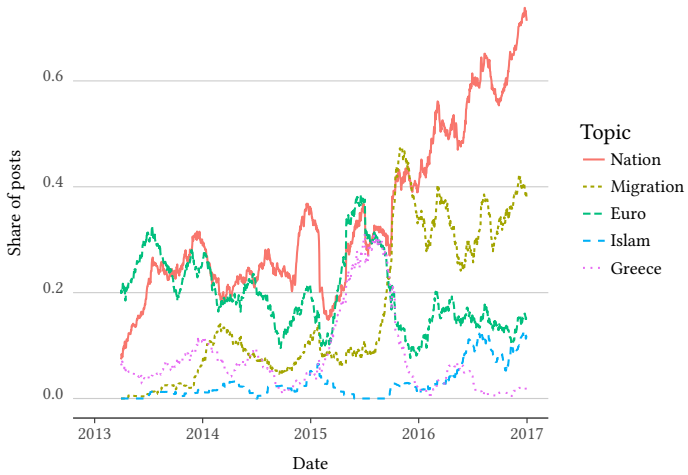


Figure 2.3. The graph shows the frequency of Facebook posts containing one of four, selected word stems/families. 90-day moving averages displayed.

radical, right-wing party) the ultra-nationalist NPD. With this body of data we can estimate a full differences-in-differences specification as follows:

$$f(\text{stem} = s)_{ipt} = \gamma_p + \delta_t + \beta \cdot \mathbb{1}\{\text{party} = \text{AfD}\} \cdot \text{Post}_t + \varepsilon_{ipt}, \tag{2.1}$$

where the dependent variable $f(\text{stem} = s)$ is the frequency (mention per 100 words) of stem s in document i (party manifesto, speech), of party p at time t. For shorter pieces of text (tweets, Facebook post), we use the following variant specification:

$$\mathbb{1}\{(\text{stem} = s) \in i\}_{ipt} = \gamma_p + \delta_t + \beta \cdot \mathbb{1}\{\text{party} = \text{AfD}\} \cdot \text{Post}_t + \varepsilon_{ipt}, \tag{2.2}$$

where $\mathbb{1}\{(\text{stem} = s) \in i\}$ is a dummy indicating whether stem s is contained in document i (tweet, post) of party p at time t. In all specifications, we include a full set of party fixed effects (γ_p) and time fixed effects (δ_t): these are year fixed effects for speeches and manifestos, and month×year fixed effects for tweets and Facebook posts. Post_t is a dummy for all periods after the Erfurt Resolution (March 2015). Standard errors

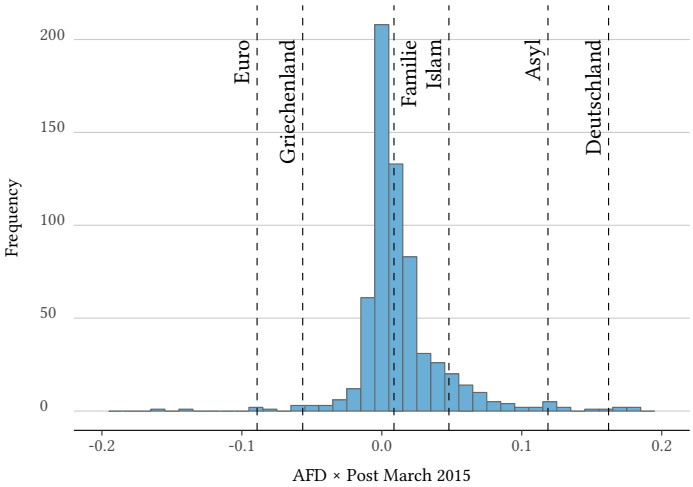


Figure 2.4. The graph shows the empirical distribution of estimated difference-in-difference coefficients, resulting from the empirical setup in equation (2.2), relating to 645 frequent word stems on Facebook, together with the location of six selected word stems.

ε_{ipt} are clustered at the party×year cell level (for manifestos and speeches) or at the party×year×month level (for tweets and Facebook posts).

The crucial difference-in-differences parameter of interest is β, indicating the increase in frequency (or mentions) of a given stem in documents of the AfD, after the Erfurt Resolution, conditional on state and time fixed effects. Table 2.1 reports the estimates of β across four media (manifestos, speeches, tweets, and Facebook posts, in Panels A through D, respectively), and for five outcome stems of interest: Greece, the Euro, Islam, migration, and nation.[21] Every cell in that table reports the estimate of the difference-in-differences parameter for one regression, defined by a dyad of medium and stem.

21. More precisely, the stem "Greece" encompasses all German words including **griech**; "Euro" all words that start with *euro**, but not *europ**, and also the acronym *EZB* (European Central Bank, in German); the stem "Islam" all words including **islam** and **muslim**; the stem "migration" all words including **migration**, **wander**, **flüchtling**, and **asyl**; the stem "nation" all words including **nation** and **deutsch**. Appendix Table B1 reports the 10 most frequent words identified by this algorithm for each stem.

Table 2.1: AfD's language change: diff-in-diff estimates

	Topic:				
	(1) Greece	(2) Euro	(3) Islam	(4) Migration	(5) Nation
Panel A: Mentions per 100 words in manifestos					
AfD × After March 2015	-0.011	-0.780***	0.052***	0.269***	-0.041
	(0.021)	(0.193)	(0.013)	(0.050)	(0.237)
Panel B: Mentions per 100 words in speeches					
AfD × After March 2015	-0.183**	-0.546***	0.063*	-0.028	0.112
	(0.070)	(0.099)	(0.034)	(0.097)	(0.100)
Panel C: Mentioned in Twitter posts					
AfD × After March 2015	-0.059***	-0.157***	0.020**	0.023**	-0.098***
	(0.009)	(0.011)	(0.009)	(0.012)	(0.018)
Panel D: Mentioned in Facebook posts					
AfD × After March 2015	-0.017	-0.055***	0.042***	0.112***	0.209***
	(0.016)	(0.021)	(0.011)	(0.023)	(0.030)

Notes: Coefficients and standard errors (in brackets) from OLS regressions. In Panel A the unit of observation is a manifesto, in Panel B a speech, in Panel C a Twitter post and in Panel D a Facebook post. All regressions include party (AFD, CDU, CSU, FDP, Grüne, Die Linke, NPD, SPD) fixed effects. Panels A and B include year fixed effects, Panels C and D month fixed effects. Number of observations: 70 (Panel A), 113 (Panel B), 66,422 (Panel C) and 40,118 (Panel D). One, two and three stars represent significance at the 10%, 5% and 1% levels respectively.

Across all text media, we see consistent results. Even when viewed in relation to the language used by the other political parties in Germany, the AfD notably reduces the mentions of Greece and the Euro in its rhetoric, and increases the usage of words related to Islam, to migration, and to Germany/the nation. For example, the estimate in Panel B, Column 2, suggests that after 2015, the reduction of mentions of stems relating to the Euro in speeches by AfD members is 0.546 per 100 words (significant at <1% level). This compares to a mean of the dependent variable of 0.703 (for AfD speeches, before 2015); it is thus a very sizeable decrease.[22]

By converse, the estimate in Panel D, Column 4, suggests that after March 2015, the share of Facebook posts mentioning a stem related to the migration context increases

22. Appendix Table B2 provides (conditional) means for all dependent variables.

by 11.2 percentage points (significant at <1% level). Again, this is sizeable if compared to a pre-March 2015 mean of the outcome variable of 5.8 percent (for the AfD).

Arguably, the five word stems shown in Table 2.1 have been arbitrarily chosen, based on our priors regarding which words should witness the starkest changes following the rightward turn imparted on the AfD after the Erfurt Resolution. To avoid our subjective bias, and to validate the stems chosen in Table 2.1, in Figure 2.4 we follow a different approach. Here, we repeat the standard differences-in-differences estimations of Equation 2.1 above, applying this regression setup to each of the 645 most frequent word stems that we identified in our entire body of Facebook posts.[23]

Figure 2.4 presents the distribution of the β coefficients estimated from Equation 2.1, across 645 stems. First, it is noticeable that the distribution of point estimates is skewed to the right of zero: this indicates that the language used by the AfD, after March 2015, becomes more varied. Second, vertical dashed lines in the figure show the positioning of the point estimates relating to key words used so far. Confirming the results of Table 2.1, we see that "Euro" and "Greece" are to the left of zero, whereas the usage of words such as "Islam", "asylum", and especially "Germany" increases dramatically for the AfD after March 2015, relative to other parties. Third, it is also noticeable that a traditional mainstay of conservative political ideology, the "family", does not move into the focus of the AfD's rhetoric: the point estimate is very close to zero. We see this as suggestive of the fact that the post-March 2015 turn experienced by the AfD was explicitly nationalistic and xenophobic (anti-Muslim), not merely conservative.

2.5 Electoral Results

2.5.1 Empirical Setup

How did the rightward turn in the AfD's rhetoric and policy platform change its electoral fortunes? We compare electoral results for the AfD before and after 2015—2015 marks a watershed, as thanks to the Erfurt Resolution and the subsequent party congress in Essen, the initial (fiscally conservative) party leaders were replaced by a new, nationalistic and xenophobic leadership (cf. the timeline of events in Figure 2.1). Specifically, we compare the results in the federal election in September 2013 to local elections to state assemblies that occurred after 2015.

In 2013, running on a strict anti-Euro platform, the AfD barely missed passing the 5% threshold to enter the federal parliament; in the period after 2015, state elections occurred in the states of Baden-Württemberg, Rhineland-Palatinate, Saxony-

23. To be more precise, we consider the universe of words in the body of Facebook posts we collected. We remove numbers, punctuation, and stopwords, and then stem the resulting words using the *tm* package for R. We keep all stems that are used at least 200 times. This results in 645 word stems.

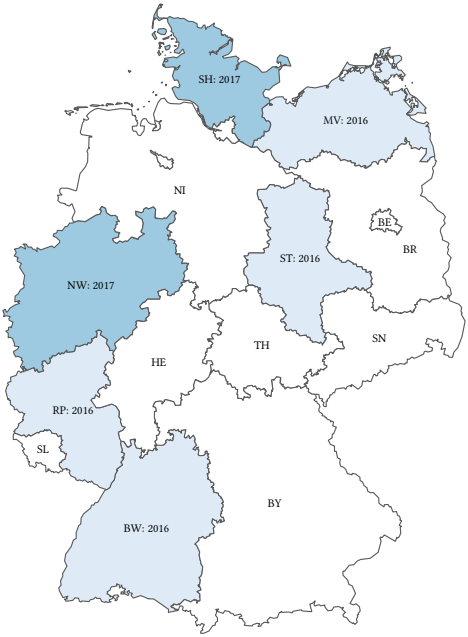


Figure 2.5. The map shows the 16 federal states in Germany, including three city states. Six states held the elections to their regional parliaments between March 2015 and September 2017 and are thus included in our empirical analyses: Baden-Württemberg (BW), Rhineland-Palatinate (RP), Saxony-Anhalt (ST), Mecklenburg-Anterior Pomerania (MV), Schleswig-Holstein (SH), and North Rhine-Westphalia (NW). The Saarland (SL) also held its regional elections in 2017; however, we disregard the Saarland as, during the Weimar era, it was occupied by French troops and no elections to the *Reichstag* were held. We also do not consider the city states of Berlin, Bremen, and Hamburg.

Anhalt, Mecklenburg-Anterior Pomerania (all in 2016), North Rhine-Westphalia, and Schleswig-Holstein (in 2017). In all of those cases, the AfD passed the 5% threshold (often reaching double-digit results) and gained seats in the state assemblies. Figure 2.5 provides a map of the states used in our analysis.[24]

Our baseline regression specification is as follows:

$$\text{ShareAfD}_{it} = \theta_s + \beta \cdot \text{ShareNazi}_i + x_{1i}'\gamma + \varepsilon_{it}, \tag{2.3}$$

where ShareAfD_{it} is the share of votes cast for the AfD in municipality i in year t (where t may either refer to the federal election of 2013, or a state election in 2016/17). Note that, in our baseline setting, we calculate the share of votes relative to *all* eligible voters, not just relative to votes cast. We do this in order to incorporate two margins of voter mobilisation towards the AfD: switching from non-voting to the AfD, or from other parties to the AfD. In a later step we will disentangle the intensive and the extensive margins of voting.

The dependent variable is regressed on a full set of state fixed effects, θ_s, a measure of support for the NSDAP party during the Weimar era, ShareNazi_i, and in some specifications also a set of (mostly time-invariant) municipal level covariates, x_{1i}, such as population or unemployment rates.

To take care of municipal-level, time-invariant omitted factors that may determine a constant inclination to vote for the AfD, the following, alternative specification takes advantage of the fact that each municipality is observed twice, once before and once after 2015, and focuses on the *change* in vote share from 2013 to 2016/17:

$$\Delta(\text{ShareAfD}_{i,2016/17-2013}) = \theta_s + \beta \cdot \text{ShareNazi}_i + x_{2i}'\gamma + \varepsilon_{it} \tag{2.4}$$

Note that, unless one assumes that time-invariant municipality characteristics have time-varying effects on the outcome variables (varying between 2013 and 2016/17), the effect of these variables will be "differenced out" in such a first-differences specification. However, one may still allow for time-varying effects of covariates, or investigate *changes* in municipal-level covariates occurring between 2013 and 2016/17. For these reasons, we may also include a vector of covariates x_{2i}, potentially different from the covariates included in Equation 2.3.

24. Note that we ignore the elections in Berlin in 2016 (as there is only one municipality in the state of Berlin) and in the Saarland in 2017 (as the Saar region did not vote for the *Reichstag* in the Weimar era, being under French occupation). Arguably, people's objectives and motives to vote in a state election may differ from a federal election, potentially confounding a comparison of voting patterns. However, to the extent that this divergence in voting behaviour affects all states in the same manner, it should not invalidate the inference we draw. In future work, we plan to analyse the electoral results of the upcoming federal election of 2017, thereby comparing two elections to the same legislative body, the *Bundestag* (2017 and 2013).

Table 2.2: AfD electoral results

	Explanatory variable (std.):			
	(1) NSDAP average	(2) NSDAP 1928	(3) NSDAP 1930	(4) NSDAP 1933
Panel A: AfD vote share 2013				
	0.0523	0.0129	0.0832*	0.0408
	(0.0443)	(0.0315)	(0.0472)	(0.0414)
Panel B: AfD vote share 2016/17				
	0.1606***	0.0943***	0.1765***	0.1226***
	(0.0372)	(0.0280)	(0.0387)	(0.0353)
Panel C: Change in AfD vote share 2013 to 2016/17				
	0.1510***	0.0945***	0.1574***	0.1149***
	(0.0318)	(0.0256)	(0.0309)	(0.0301)
State FEs	✓	✓	✓	✓
Observations	5861	5861	5861	5861

Notes: The Column header indicates the respective explanatory variable used, the Panel header the respective dependent variable. The explanatory variable in Column (1) is the standardised average of NSDAP vote shares in 1928, 1930, and 1933. All variables (explanatory and dependent) are standardised. Sample includes municipalities in the six German states of Baden-Württemberg, Mecklenburg-Anterior Pomerania, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt and Schleswig-Holstein. All regressions include state fixed effects. Standard errors are clustered at the level of historic municipalities/counties. One, two and three stars represent significance at the 10%, 5%, and 1% levels respectively.

Of the 5,475 municipalities in the six states considered, 1,191 are "exactly matched": that is, they are either matched to a *Stadtkreis* (city-county) of the Weimar era (84 cases), or to a municipality contained in a larger county, but which had more than 2,000 inhabitants in the Weimar era (1,107 cases), thus with exact electoral returns in the 1920s and 30s. The remaining 4,284 municipalities are assigned one of 259 Weimar-era "remainders of a county". To account for potential correlation between these multiple observations assigned to a single historical electoral result, we cluster all error terms ε_{it} at the Weimar-era unit of observation (*Stadtkreis*, municipality above 2,000 inhabitants, or "remainder of the county").

2.5.2 Baseline Results

Table 2.2 reports our baseline estimates. To facilitate the interpretation of the coefficients, all variables, dependent and explanatory, are standardised. The four columns represent four possible, alternative explanatory variables: the NSDAP vote share in 1928, 1930, and 1933 in columns 2–4, respectively,[25] and an aggregate z-score index (also standardised) of these three elections in Column 1. The regressions in Panel A feature the AfD's vote share in the federal election of 2013 as the dependent variable. Across all columns (all explanatory variables), one can see that there is a positive, but generally small and insignificant correlation between Nazi vote in the Weimar era and AfD support.

However, when one looks at support for the AfD in the state elections occurring in 2016/17, after the Erfurt resolution, the results are very different. As one can see in Panel B, the correlation between past Nazi support and contemporary AfD support, in elections in which the AfD represented a far right, xenophobic platform, is strong and significant. In our preferred specification of Column 1, a one standard deviation increase in NSDAP votes in the Weimar era corresponds to a 0.16 standard deviations higher vote share for the AfD. Notably, the results are very similar no matter whether one uses an early indicator of Nazi support, such as votes in 1928 (when the NSDAP was still a fringe party, obtaining only 2.6 percent of the votes at the national level), or a late indicator, such as the last, semi-free election of March 1933, when the NSDAP obtained 43.9 percent of the votes.

Finally, the results in Panel C correspond to the regression setup in Equation 2.4. Here, by looking at changes in vote shares, we take into account municipality-specific unobservables that may determine a municipality's inclination to vote for a right-wing alternative to the established parties such as the CDU, no matter whether the party is mostly fiscally conservative (as the AfD in 2013), or mostly nationalistic

25. In 1924, the NSDAP did not present a separate list for the *Reichstag* election, but supported the *Deutschvölkische Freiheitspartei* (DVFP). The electoral results of 1932 were not published at the disaggregate level.

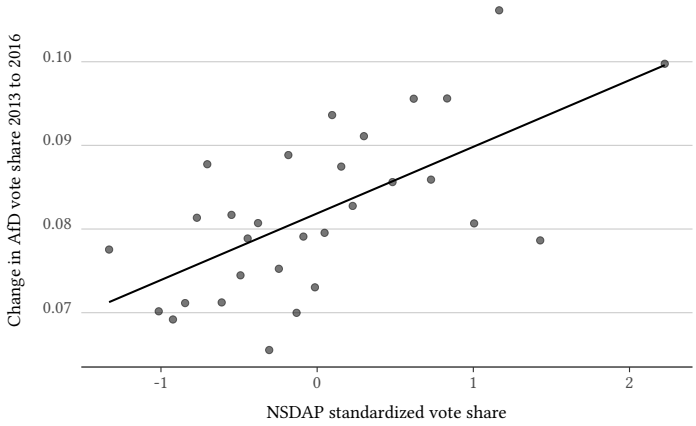


Figure 2.6. The graph shows the relationship, residualised on state Fixed Effects, between historical Nazi vote share and the change in votes for the AfD between 2013 and 2016/17 as a binned scatterplot.

and anti-immigrant (as the AfD in 2016/17). The resulting picture is very similar to Panel B. The standardised beta coefficient on the association between Nazi vote and the shift to the AfD, following its 2015 turn, is about 0.15. The estimated coefficient is sizeable compared to other, plausible determinants of far right electoral outcomes: for example, trade exposure is associated with votes for the far right in France with a beta coefficient of 0.07 (Malgouyres, 2017), or with a beta coefficient of 0.28 in Germany (Dippel, Gold, and Heblich, 2016).[26]

The seminal work by Voigtländer and Voth (2012, 2015) has shown how antisemitism is a persistent feature of certain regions in Germany. In Table 2.3 we argue, however, that antisemitism is not what explains the success of the AfD in more recent

26. In the Appendix, we examine the robustness of these baseline results as follows. First, Table B3 limits the analysis to the 1,239 municipalities that are exactly matched between today and the Weimar-era data (recall that electoral returns for municipalities with less than 2,000 inhabitants were not published in the 1920s and 30s, and that we thus matched these municipalities with the votes cast in the "Remainder of the county"). Second, Table B4 weights the regressions by a municipality's population size in 2015. In both cases, the standardised coefficients are now smaller (by about half), suggesting that the strongest historical persistence occurs in smaller municipalities. Finally, Table B5 limits the analysis to the set of exactly matched municipalities, and weights regressions by current population size.

years. Columns 1 and 2 analyse how the AfD's electoral results are correlated with the electoral success of two parties standing for the *Reichstag* election of 1924: the *Deutschnationale Volkspartei* (DNVP) and the *Deutschvölkische Freiheitspartei* (DVFP). The DNVP was the main conservative party of the Weimar era, before the emergence of the NSDAP: nationalist, reactionary, monarchist. The DVFP was split off the DNVP, as some of its members thought it should be more explicitly antisemitic. The 1924 election thus pitted two far right parties against each other: a staunchly conservative one (the DNVP), and a clearly antisemitic one (the DVFP). As the results in Table 2.3 show, the electoral success of the AfD is highly correlated with the conservative party in the Weimar era, but not with its antisemitic spin-off.

Columns 3–5 of Table 2.3 confirm these results. Here, we limit the analysis to the 423 cities that are featured both in our dataset and in the work by Voigtländer and Voth (2012). In Column 3, we first confirm that our baseline estimate of Table 2.2 can be replicated, with broadly similar results within those 423 cities. In Column 4 we then regress the AfD's electoral fortunes on the composite measure created by Voigtländer and Voth (2012): a z-score index encompassing six measures of antisemitism in the 1920s and 30s.[27] There is virtually zero (or even a negative) correlation between these expressions of early 20th-century antisemitism and AfD support, both in 2013 and in 2016/17. Finally, in Column 5 we use the indicator variable for the occurrence of pogroms in the wake of the Black Death of 1348 (again, as collected by). If anything, the correlation between mediaeval Jewish hatred —which has been shown to be a consistent predictor Nazi support—and AfD support is negative.

These findings suggest that what persisted between the Weimar era and today, and determines the AfD's electoral success, is not antisemitism but rather a right-wing ideology. In fact, the AfD was successful at keeping antisemitism out of its official policy platforms and actually explicitly endorsing Israel;[28] its religious animus is clearly more directed against Islam. Rather, the common ground between the NSDAP and the AfD in its post-2015 incarnation is nationalism and a closure towards all things foreign, especially as a reaction to economic distress.

2.5.3 Robustness Checks

The results presented so far were simple bivariate correlations, conditional only on state fixed effects. In the following, we examine the robustness to the inclusion of plausible determinants of electoral behaviour: both historical (variables that may explain the predominance of NSDAP voters in the 1920s and 30s) and contemporary

27. This index included measures for: pogroms in the 1920s, the share of DVFP votes 1924, the share of NSDAP votes 1928, letters to the Stürmer (an antisemitic newspaper), deportations per 100 Jews in 1933, and an indicator variable for whether a synagogue was destroyed (or damaged).

28. At the same time, however, several elected officials of the AfD (especially in Baden-Württemberg's state legislature) have expressed antisemitic attitudes.

Table 2.3: AfD electoral results: persistence of antisemitism

	Explanatory variable (std.):				
	(1) DNVP 1924	(2) DVFP 1924	(3) NSDAP average	(4) Anti-semitism	(5) Black death pogroms
Panel A: AfD vote share 2013					
	0.0559	0.0140	0.0401	-0.0510**	-0.0982
	(0.0473)	(0.0290)	(0.0369)	(0.0246)	(0.0609)
Panel B: AfD vote share 2016					
	0.2392***	0.0119	0.1135***	-0.0178	-0.1264**
	(0.0433)	(0.0350)	(0.0327)	(0.0247)	(0.0495)
Panel C: Change in AfD vote share 2013 to 2016					
	0.2322***	0.0078	0.1057***	-0.0018	-0.0999**
	(0.0373)	(0.0331)	(0.0313)	(0.0244)	(0.0500)
State FEs	✓	✓	✓	✓	✓
Observations	5860	5860	423	423	423

Notes: The Column header indicates the respective explanatory variable used, the Panel header the respective dependent variable. The explanatory variable in Column (4) is the standardised first principal component of six measures of 1920s/30s antisemitism, as in Voigtländer and Voth (2012). All variables (explanatory and dependent) are standardised, except the indicator variable for Black Death Pogroms in Column (5), which has a mean of 0.251. Sample includes municipalities in the six German states of Baden-Württemberg, Mecklenburg-Anterior Pomerania, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt and Schleswig-Holstein. All regressions include state fixed effects. Standard errors are clustered at the level of historic municipalities/counties. One, two and three stars represent significance at the 10%, 5%, and 1% levels respectively.

(present-day sociodemographics as correlates of electoral outcomes). In Table 2.4, we examine how our preferred specification of Table 2.2, Panel C, using the change in AfD votes from 2013 to 2016/17 as the dependent variable, is sensitive to the inclusion of these covariates. Column 1 of Table 2.4 first presents the baseline estimate (without controls) as a benchmark.

In the following columns, we add control variables related to population, religion, and employment structures. In Panel A, we only include the controls relating to the Weimar era. In Panel B, we only include the controls relating to the present day. Finally, in Panel C we repeat each regression including both historical and contemporary controls. Starting in Column 2, we consider the domain of "population": we control either for the (log) size of the municipality in the 1920s/30s, or for the current (log) size of the municipality and for an urbanisation category dummy, or for all of these variables together. In neither case is the baseline estimate modified substantially.

When we consider the domain of "religion", in Column 3, we control for the population shares of Catholics and Jews in 1925 in Panel A (the omitted category is Protestants and "others", the latter being negligible in 1925), and for the population shares of Catholics and "others" (including Muslims, other religions, and atheists) in Panel B (the omitted category is Protestants). The inclusion of this set of controls, no matter whether contemporary or historical, changes the magnitude (but not the precision) of the estimated coefficients quite substantially: the estimated beta coefficients drops by about half from 0.15 to 0.06–0.07. The crucial explanatory factor here is the presence of Catholics: as pointed out by a large literature, most recently by Spenkuch and Tillmann (2017), Catholic regions were, ceteris paribus, less likely to vote for the NSDAP. Our analysis shows that this holds also for today's support for the AfD, even in a within-state setting. Nevertheless, albeit dampened, the correlation between Nazi support and AfD electorate today remains quite substantial and highly significant.

Column 4 considers another major determinant of voting behaviour: the economic/-social structure, and the economic conditions (especially distress caused by unemployment). In Panel A, we control for the historical employment structure in municipalities or counties: shares of employed in industry/manufacturing, in commerce, and in administration (agriculture and other sectors being the omitted category). We also control for unemployment rates in 1933, at the peak of the Great Depression in Germany. In Panel B, we control for the official county-level unemployment rate in 2015. Across all Panels, including controls for the employment structure does not affect the baseline correlation between historical Nazi support and contemporary votes for the AfD (if anything, the correlation becomes stronger).

Finally, Column 5 pools all control variables from columns 2–4 together. Even in this most demanding specification, the correlation between NSDAP support and support for the AfD remains strong and significant: the estimated beta coefficient in Panel C,

Table 2.4: AfD electoral results including controls

	Dependent variable:				
	Change in AfD vote share, 2013 to 2016 (std.)				
	(1) Baseline	(2) Population	(3) Religion	(4) Employment	(5) Full
Panel A: With historical controls					
NSDAP share (std.)	0.1510***	0.1511***	0.0652**	0.1931***	0.0835**
	(0.0318)	(0.0318)	(0.0285)	(0.0441)	(0.0385)
Observations	5861	5855	5407	4520	4359
Panel B: With contemporary controls					
NSDAP share (std.)	0.1510***	0.1512***	0.0705***	0.1418***	0.0654**
	(0.0318)	(0.0317)	(0.0268)	(0.0307)	(0.0264)
Observations	5861	5861	5088	5827	5057
Panel C: With historical and contemporary controls					
NSDAP share (std.)	0.1510***	0.1516***	0.0575**	0.1795***	0.0698*
	(0.0318)	(0.0316)	(0.0280)	(0.0433)	(0.0382)
Observations	5861	5855	4998	4487	4008
State FEs	✓	✓	✓	✓	✓

Notes: The dependent variable is the change in vote share for the AfD (relative to eligible voters) from 2013 to 2016/17. Each Column adds a different set of historical control variables. The explanatory variable across all columns is the average NSDAP vote share across 1928, 1930, and 1933 (standardised). Population controls are: [historical] log population size (average of 1925 and 1933); [contemporary] log population size in 2015 and urbanisation code dummies (3 categories). Religion controls are: [historical] the share of Catholics and Jews; [contemporary] the share of Catholics and "Others" (i.e., Muslims, other religions, and no religion). Employment controls are: [historical] shares of employed in industry and manufacturing, employed in trade and commerce, and employed in administration (agriculture and "other sectors" is the omitted category), all measured in 1925, as well as the unemployment share in 1933; [contemporary] the unemployment rate in 2015. Sample includes municipalities in the six German states of Baden-Württemberg, Mecklenburg-Anterior Pomerania, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt and Schleswig-Holstein. All regressions include state fixed effects. Standard errors are clustered at the level of historic municipalities/counties. One, two and three stars represent significance at the 10%, 5%, and 1% levels respectively.

when both historical and contemporary controls are included, is approximately 0.07.

Table 2.5 looks more closely at the spatial heterogeneity of results, and at current determinants of voting that may explain the shift towards the AfD from 2013 to 2016/17. After reproducing the baseline estimate again in Column 1, in Column 2 we investigate whether the effect of historical NSDAP voting differs between states East and West Germany. Radical right movements (especially neo-Nazis) and xenophobia have long been a problem in the states of formerly communist East Germany. However, even if generally support for the AfD (post-2015) is higher in the East (in our sample of elections, 10.7% in the West and 24.0% in the East), the pattern of historical persistence is nearly identical. The two coefficients suggest a standardised correlation ranging between 0.11 in the East and 0.15 in the West; a test for inequality of coefficients yields a p-value of 0.78.

Arguably the most important political event in Germany in 2015 was the sudden and dramatic influx of refugees, mostly fleeing the Syrian civil war. Large numbers of them—hundreds of thousands—reached Germany on foot, via the Balkans and Austria, after Germany's decision, in September 2015, to suspend the Dublin agreement and not to limit their intake. The refugees were allocated to states and counties according to their size and GDP; however, within counties, the allocation of refugees to municipalities was idiosyncratic. The effect of the refugee inflow on votes for the far right is ambiguous. On the one hand, refugees are often perceived as a threat and a potential source of crime, moving voters to the right (Dustmann, Vasiljeva, and Piil Damm, 2016); on the other hand, in line with Allport's (1954) "contact hypotheses", direct acquaintance with refugees can actually increase empathy and support for moderate parties (Steinmayr, 2017).

In Column 3, we control for the presence of refugees in each municipality (calculated as a share relative to total population, as of late 2016). The effect is negative, suggesting that more refugees lead to *fewer* votes for the AfD, and negligible: the standard deviation of the "share refugees" variable is 0.004. Increasing the share of refugees by one standard deviation thus decreases the vote share of the AfD by less than half a percentage point of a standard deviation.

Table 2.5: AfD electoral results: local economic shocks

	Dependent variable: change in AfD vote share 2013 to 2016 (std.)					
	(1) Baseline	(2) East vs. West	(3) Refugees	(4) Unempl.	(5) Both	(6) With Controls
NSDAP share (std.)	0.1510***		0.1508***	0.1439***	0.1436***	0.0665*
	(0.0318)		(0.0318)	(0.0306)	(0.0307)	(0.0350)
NSDAP share (std.) × East		0.1144				
		(0.1331)				
NSDAP share (std.) × West		0.1526***				
		(0.0327)				
Share refugees			-1.1094		-1.4516	-1.2306
			(2.4535)		(2.4239)	(2.9898)
% change unemployed (2013–16)				0.0497*	0.0518*	0.0880***
				(0.0293)	(0.0291)	(0.0337)
State FEs	✓	✓	✓	✓	✓	✓
Historical controls						✓
Contemporary controls						✓
Observations	5861	5861	5861	5768	5768	3949

Notes: The dependent variable is the change in vote share for the AfD (relative to eligible voters) from 2013 to 2016/17. NSDAP share is defined as the standardised average of vote shares in 1928, 1930, and 1933. Sample includes municipalities in the six German states of Baden-Württemberg, Mecklenburg-Anterior Pomerania, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt and Schleswig-Holstein. All regressions include state fixed effects. Standard errors are clustered at the level of historic municipalities/counties. One, two and three stars represent significance at the 10%, 5%, and 1% levels respectively.

Globalisation, the decline of manufacturing, and a decrease in job security are often cited as a cause of the far right's recent electoral fortunes. Overall, Germany had a comparatively strong economy in the time frame considered (2013 to 2016/17), and among developed countries it remains among those with the highest shares of employment in (skilled) manufacturing, and the lowest rates of unemployment, also among youths. In fact, across the municipalities in our dataset, between 2013 and 2016 the number of unemployed individuals decreased by 7 percent on average. In Column 4, we control for the change in unemployment: as predicted, an increase in unemployment leads to a larger change in the AfD's vote share between 2013 and 2016/17. However, the effect is comparatively small and leaves the main coefficient of interest virtually unaffected. Including both the controls for refugee presence and for unemployment (Column 5) also does not affect the historical persistence of NSDAP voting; when all controls, historical and contemporary, of Table 2.4 are additionally included, the results are still stable (Column 6).

2.5.4 Voter Migration

Recall that, in all regressions shown so far, the dependent variable was defined as the share of votes cast for the AfD relative to the eligible voting population: the goal was to encompass both margins of voter mobilisation towards a party moving to the far right. In what follows, we want to understand the importance of the intensive margin (voters moving from other parties to the AfD) and the extensive margin (voter mobilisation from non-voting to voting).

We begin with the analysis of turnout. Table 2.6, Column 1, presents regressions in which the dependent variable is the voter turnout in the elections of 2013 and 2016/17 (in Panels A and B, respectively), and the change in voter turnout between 2013 and 2016/17 (Panel C). Higher Nazi support in the 1920s and 30s is associated with moderately lower turnout in 2013: a one standard deviation higher NSDAP support translates into half a percentage point lower turnout, relative to a sample mean of almost 64 percent. In 2016, the association turns positive, but is small and insignificant. More interestingly, between 2013 and 2016/17 voter turnout increased by 3.6 percentage points overall, and this increase is substantially and significantly correlated with historical Nazi vote: a one standard deviation higher NSDAP support is associated with a 0.7 percentage points larger increase in turnout.

Having established that the mobilisation of non-voters is strongly associated with historical Nazi support, we now look at the intensive margin: conditional on voting, which parties are chosen? We first, in Column 2, look at the AfD. As opposed to the previous tables, we now use the (unstandardised) AfD vote share *relative to votes cast* as the dependent variable, rather than relative to eligible voters. The picture, however, is similar. There is a small and insignificant correlation of historical Nazi support with

Table 2.6: Persistence of voting: turnout and other parties

	Dependent variable:					
	(1) Turnout	(2) AfD	(3) CDU	(4) SPD	(5) NPD	(6) Linke
Panel A: 2013						
NSDAP share (std.)	-0.0053**	0.0015	-0.0238***	0.0160***	0.0017***	0.0033***
	(0.0021)	(0.0010)	(0.0069)	(0.0047)	(0.0005)	(0.0011)
Mean dep. var.	0.6385	0.0473	0.4628	0.2410	0.0144	0.0751
Panel B: 2016						
NSDAP share (std.)	0.0021	0.0090***	-0.0212***	0.0120***	0.0012***	0.0003
	(0.0023)	(0.0022)	(0.0058)	(0.0039)	(0.0004)	(0.0004)
Mean dep. var.	0.6748	0.1291	0.3260	0.2688	0.0112	0.0415
Panel C: Change 2013 to 2016						
NSDAP share (std.)	0.0074***	0.0075***	0.0026	-0.0039	-0.0013***	-0.0030***
	(0.0023)	(0.0017)	(0.0026)	(0.0027)	(0.0004)	(0.0008)
Mean dep. var.	0.0363	0.0818	-0.1368	0.0277	-0.0051	-0.0336
State FEs	✓	✓	✓	✓	✓	✓
Observations	5861	5861	5861	5861	4725	5861

Notes: The dependent variable is turnout (total votes cast relative to eligible voters) in Column 1; the vote share (relative to total votes cast) of AfD, CDU, SPD, NPD and Linke, respectively, in Columns 2–6. NSDAP share is defined as the standardised average of vote shares in 1928, 1930, and 1933. Sample includes municipalities in the six German states of Baden-Württemberg, Mecklenburg-Anterior Pomerania, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt and Schleswig-Holstein. All regressions include state fixed effects. Standard errors are clustered at the level of historic municipalities/counties. One, two and three stars represent significance at the 10%, 5%, and 1% levels respectively.

AfD votes in 2013, and a positive, strong and significant correlation with AfD votes in 2016, and with the change in votes cast for the AfD.

In columns 3 and 4 we consider the two largest parties in Germany, the moderately conservative CDU and the social democratic SPD. The associations of historical Nazi support with votes for these two parties are each other's mirror image: CDU voters are, on average, located where historically the NSDAP was less represented, whereas SPD voters are more likely to be located in NSDAP strongholds. This historical association can plausibly be explained by sociodemographic characteristics (urban vs. rural, presence of Catholics); importantly, however, the comparison of Panels A and B (or, the results in Panel C) shows that this correlation does not change over time, even as the electoral fortunes of these two parties vary.

Instead, the results of the far-right NPD are consistently positively correlated with the NSDAP vote (Column 5). The implied beta coefficient is approximately 0.1, thus slightly lower than the correlation with the AfD vote in 2016/17. Looking at Panel C, one can see that higher Nazi vote shares in the past are associated with a move away from the NPD between 2013 and 2016/17, suggesting that this is a potential electoral basin the AfD drew from—albeit one should obviously be cautious with inferences on individual voter migration made based on aggregate data. A similar picture emerges for the "Linke" party, the far-left option in the German political spectrum (Column 6). In 2013, there is a positive correlation between historical Nazi vote and contemporary Linke vote shares; this correlation all but disappears by 2016/17. As was the case for the NPD, the change in votes going to the Linke is negatively associated with historical Nazi support.

Again, one should caution against drawing conclusions on individual voter migration based on aggregate data. The historical correlations in Table 2.6 are, however, at least consistent with a world in which the electoral success of the AfD post-2015 draws from two sources: from mobilising former non-voters, and from former voters of other extreme parties, both on the far left and the far right end of the political spectrum.[29]

2.6 Conclusion

We have argued that a hitherto unexplored historical persistence of right-wing ideology is an important determinant of recent electoral outcomes in Germany. After 2015 an existing party, the 'Alternative für Deutschland', abruptly repositioned itself at the far-right of the political spectrum and started espousing nationalist, openly xenophobic views. We document this change by analysing the content of media (speeches, manifestos, Twitter and Facebook posts) produced by all parties from 2008

29. This is also consistent with the survey-based analysis of voter migration flows presented in Hensel et al. (2017, Table 1).

onwards.

As a result of this shift of platform, the AfD benefited from an underlying but hitherto unrevealed persistence: municipalities that supported the NSDAP during the Weimar Republic voted proportionally more for the AfD in the 2016/17 state elections. We showed that this relationship is robust to conditioning on state fixed effects and a wide range of variables that might plausibly be correlated with right-wing voting. Importantly, the relationship holds when controlling for the local inflow of Syrian refugees and for local unemployment shocks.

It remains for further research to better understand the underlying mechanisms leading to this persistence. Given other work in this area showing high correlations between parents' and childrens' support for far-right parties (Avdeenko and Siedler, 2017), it is likely that this persistence reflects the vertical transmission of preferences coupled with a relatively immobile population. In the future, longitudinal survey data on AfD survey supporters will allow us to test this hypothesis more directly.

Appendix B.1: Matching contemporary and historical election data

We match present-day electoral outcomes to the Weimar era party support in two steps:

Step 1: First, we identify the boundary of each county with electoral data in the Falter and Hänisch (1990) dataset, using the county name to match counties to polygons in the shapefile provided by the Census Mosaic project[30]. We then identify coordinates for each Weimar era municipality (to the best of our knowledge, no shapefiles of municipalities in the period are available): for each municipality, we first use OpenStreetMaps' Nominatim API to search for modern administrative centers, villages, towns, cities or suburbs sharing a name with the historic municipality. We overlay the returned coordinates on the county map and discard any results which lie outside the boundary of the county to which the historic municipality belongs, according to the Falter and Hänisch (1990) data. In this way, we obtain valid latitude and longitude coordinates for around two thirds of the Weimar era municipalities. For municipalities which return no valid matches, for example because of name changes between the Weimar era and today, we manually search for coordinates. To do so, we use a combination of sources including `http://gov.genealogy.net`, a database of historic geographies, and Wikipedia. We check the manual lookups for validity by ensuring that the coordinates lie within the boundaries of the county to which the municipality belongs, again according to the Falter and Hänisch (1990) data.

Step 2: In this step, we match contemporary municipalities to a Weimar era geography for which the Falter and Hänisch (1990) dataset provides electoral data. If a modern municaplity's coordinates (provided by DESTATIS) are within 2.5 kilometers of the coordinates of a municipality identified in Step 1, we match the contemporary district to the electoral data from that historic municipality. Otherwise, we overlay the coordinates of the modern municipality on top of the shapefile of counties and assign the electoral results for the "remainder of the county" to the modern municipality. Because electoral geography is not constant between 1924 and 1932, a modern municipality can be matched to different entities for different election years.

30. Electoral geography changes between the years 1924 and 1932, the result of counties being merged or split and other boundary changes. We thus match counties to boundaries seperately for each of the 1924, 1928, 1930 and 1932 elections. In a very small number of cases, we make changes to the county shapefiles in order to better match the county/municipality hierarchy provided by the Falter and Hänisch (1990) dataset.

Appendix B.2: Extra Figures and Tables

Figure B1. AfD Electoral poster for the federal election of September 2013. It reads: "Greeks are desperate. Germans are paying. Banks are cashing in. Stop this."

Figure B2. AfD Electoral poster for the state election in Baden-Württemberg in March 2016. It reads: "For our state – for our values. Immigration needs clear rules."

Table B1: Most frequent words (stems in Table 2.1)

Greece	Euro	Islam	Migration	Nation
griechenland	euro	islamischen	zuwanderung	deutschlands
griechische	eurokrise	muslime	flüchtlingen	deutsche
griechischen	euroraum	islam	asyl	deutscher
griechen	ezb	islamistische	migration	deutsch
griechenlands	eurostaaten	islamistischen	einwanderung	nationale
grieche	eurozone	islamische	flüchtlingspolitik	national
griechisch	eurorettung	islamisten	asylverfahren	nationalen
griechenlandanleihen	euros	islamischer	zuwanderer	deutschen
griechischer	eurobonds	muslimen	asylsuchende	deutsches
griechenlandkrise	eurojust	muslimischen	einwanderer	nationaler

Table B2: AfD's language change: means of dependent variables (Table 2.1)

	(1) Greece	(2) Euro	(3) Islam	(4) Migration	(5) Nation
Panel A: Mentions per 100 words in manifestos					
Mean (overall)	0.005	0.140	0.030	0.264	0.610
Mean (all parties, pre-March 2015)	0.013	0.292	0.013	0.170	1.028
Mean (AfD)	0.015	0.401	0.046	0.405	0.963
Mean (AfD, pre-March 2015)	0.031	1.043	0.000	0.199	1.406
Panel B: Mentions per 100 words in speeches					
Mean (overall)	0.055	0.145	0.033	0.126	0.556
Mean (all parties, pre-March 2015)	0.088	0.235	0.023	0.074	0.498
Mean (AfD)	0.102	0.367	0.028	0.128	0.789
Mean (AfD, pre-March 2015)	0.216	0.703	0.000	0.105	0.697
Panel C: Mentioned in Twitter posts					
Mean (overall)	0.011	0.021	0.010	0.043	0.086
Mean (all parties, pre-March 2015)	0.012	0.027	0.009	0.022	0.089
Mean (AfD)	0.053	0.133	0.009	0.027	0.113
Mean (AfD, pre-March 2015)	0.069	0.175	0.002	0.016	0.133
Panel D: Mentioned in Facebook posts					
Mean (overall)	0.019	0.055	0.024	0.088	0.230
Mean (all parties, pre-March 2015)	0.017	0.059	0.017	0.040	0.200
Mean (AfD)	0.064	0.184	0.044	0.166	0.371
Mean (AfD, pre-March 2015)	0.068	0.214	0.015	0.058	0.231

Notes: Table reports means for five groups of words. These are the dependent variables in the diff-in-diff regressions of Table 2.1. Overall means (first row in each panel) and conditional means reported.

Table B3: AfD electoral results exact matches: 2013, 2016/17

	Explanatory variable (std.):			
	(1) NSDAP average	(2) NSDAP 1928	(3) NSDAP 1930	(4) NSDAP 1933
Panel A: AfD vote share 2013				
	0.0004	-0.0245	0.0487*	-0.0021
	(0.0249)	(0.0203)	(0.0267)	(0.0234)
Panel B: AfD vote share 2016				
	0.0558***	0.0337**	0.0411**	0.0622***
	(0.0179)	(0.0158)	(0.0178)	(0.0205)
Panel C: Change in AfD vote share 2013 to 2016				
	0.0584***	0.0434***	0.0269*	0.0659***
	(0.0160)	(0.0143)	(0.0159)	(0.0196)
State FEs	✓	✓	✓	✓
Observations	1239	1239	1239	1239

Notes: Sample includes municipalities in the six German states of Baden-Wuertemberg, Mecklenburg-West Pomerania, Rhine-Palatine, Saxony-Anhalt and Schleswig-Holstein. Controls include the unemployment rate in 2015, size of area, population in 2014, percent of population male, and urbanisation status. Standard errors are clustered on the historic district to which modern districts are matched. One, two and three stars represent significance at the 10%, 5% and 1% levels respectively.

Table B4: AfD electoral results weighted: 2013, 2016/17

	Explanatory variable (std.):			
	(1) NSDAP average	(2) NSDAP 1928	(3) NSDAP 1930	(4) NSDAP 1933
Panel A: AfD vote share 2013				
	0.0616*	0.0207	0.0643	0.0554
	(0.0334)	(0.0254)	(0.0391)	(0.0404)
Panel B: AfD vote share 2016				
	0.0995***	0.0730***	0.0645***	0.0897***
	(0.0203)	(0.0168)	(0.0219)	(0.0208)
Panel C: Change in AfD vote share 2013 to 2016				
	0.0838***	0.0696***	0.0463**	0.0757***
	(0.0194)	(0.0194)	(0.0224)	(0.0177)
State FEs	✓	✓	✓	✓
Observations	5861	5861	5861	5861

Notes: Sample includes municipalities in the six German states of Baden-Wuertemberg, Mecklenburg-West Pomerania, Rhine-Palatine, Saxony-Anhalt and Schleswig-Holstein. Controls include the unemployment rate in 2015, size of area, population in 2014, percent of population male, and urbanisation status. Standard errors are clustered on the historic district to which modern districts are matched. One, two and three stars represent significance at the 10%, 5% and 1% levels respectively.

Table B5: AfD electoral results exact matches and weighted: 2013, 2016/17

	Explanatory variable (std.):			
	(1) NSDAP average	(2) NSDAP 1928	(3) NSDAP 1930	(4) NSDAP 1933
Panel A: AfD vote share 2013				
	0.0695	0.0019	0.0829	0.0785*
	(0.0435)	(0.0327)	(0.0548)	(0.0468)
Panel B: AfD vote share 2016				
	0.0777***	0.0594***	0.0523**	0.0718***
	(0.0246)	(0.0193)	(0.0264)	(0.0262)
Panel C: Change in AfD vote share 2013 to 2016				
	0.0584**	0.0616***	0.0273	0.0493**
	(0.0236)	(0.0234)	(0.0267)	(0.0223)
State FEs	✓	✓	✓	✓
Observations	1239	1239	1239	1239

Notes: Sample includes municipalities in the six German states of Baden-Wuertemberg, Mecklenburg-West Pomerania, Rhine-Palatine, Saxony-Anhalt and Schleswig-Holstein. Controls include the unemployment rate in 2015, size of area, population in 2014, percent of population male, and urbanisation status. Standard errors are clustered on the historic district to which modern districts are matched. One, two and three stars represent significance at the 10%, 5% and 1% levels respectively.

3

DISCRIMINATION EVERYWHERE? EVIDENCE FROM NEW YORK TAXIS

3.1 Introduction

Across a range of settings, non-whites suffer worse market outcomes than whites. The black-white wage differential in the U.S. labour market, adjusted for observable factors such as education and experience, is close to 15% (Wilson and Rodgers, 2016). Audit studies have shown that job applicants with distinctively non-white names are significantly less likely to receive callbacks for interviews (Bertrand and Mullainathan, 2004; Jacquemet and Yannelis, 2012). Whites receive preferred treatment in the sharing economy (Edelman, Luca, and Svirsky, 2017), on online platforms (Pope and Sydnor, 2011) and when standing for political office (Stephens-Davidowitz, 2014). An enduring question is whether these outcomes result from racial animus (*taste-based discrimination*) or from the use of racial status as an informative signal about other characteristics (*statistical discrimination*). In most non-experimental settings these are observationally equivalent, so it has been hard to isolate these motives from one another.[1]

In this paper, I analyse the differences in tips received by white and non-white drivers of taxis in New York. The setting is uniquely informative about consumers' preferences towards non-whites. In this setting there is no rational for statistical discrimination, but we can expect any anti-minority *preferences* of customers to be reflected in a tip differential: customers face no monetary cost from expressing discriminatory

1. Bertrand and Duflo (2017) provide a review of experimental and field studies aimed at identifying discrimination. One of the few studies to exploit a setting without motives for statistical discrimination is that of Soltas and Broockman (2017), who exploit a natural experiment in Illinois to show evidence for taste-based discrimination against non-white political candidates.

preferences. These discriminatory preferences might come in the form of malice towards drivers of a particular group, or because of differences in levels of altruism towards different groups. Saunders and Lynn (2010) show that altruism towards service providers is a major motive for tipping, and a large body of evidence has shown that individuals are less altruistic towards outgroups (Ockenfels and Werner, 2014).[2] Given the low stakes and high frequency of the transaction, implicit bias can also be expected to be reflected in tipping behaviour. In addition, tips are typically not observed by any third parties, so passengers have no signalling motive. Finally, given the number of passengers and drivers in New York, the probability of repeated interaction with a driver is very low, reducing the role of reputational concerns.

My analysis use ride-level administrative data from the New York Taxi and Livery Company (T&LC) on all rides made in New York City 'yellow' cabs in 2009 and 2013. The data includes a large number of journey characteristics including, for rides paid by credit card, the amount tipped. In addition, a driver identifier can be matched to the T&LC's data on registered taxi drivers. The register does not include any data on race or ethnicity, so I predict these using a novel surname-based procedure. This uses data from a genealogical website on the global distribution of surnames as well as U.S. census data and aggregate statistics from the T&LC to calculate the probability that a driver with a given surname driver belongs to a given racial or ethnic group.

Having done so, I show that non-white drivers receive tips which are around one to two percent less than those of white drivers. I demonstrate that the differential is not sensitive to controlling for selection of drivers into different types of journeys, into different vehicles or into matches with different types of passengers. Using a broad range of proxies for service quality, I find no evidence that the differential reflects variation in the quality of service by drivers of different racial categories. Although I can not completely rule out that the differential stems from unobserved heterogeneity in the service levels provided by drivers, the evidence is suggestive of some preference-based discrimination. However, the level is much lower than might be expected. Considering journeys which are identical regards to fare amount, start location, end location and time of day, white drivers receive higher tips than non-white drivers 38% of the time, whereas non-white drivers receive higher tips in 36% of cases. That is, the differential might be eliminated if one journey out of a hundred was tipped differently. What difference does exist is largely driven by passengers making more 'generous' tips—that is, tips above the social norm—to white drivers. This is consistent with evidence that in-group favouritism, rather than out-group malice, explains some discriminatory outcomes (Greenwald and Pettigrew, 2014).

2. Tipping has similarities with the dictator game, a mainstay of experimental economics. In this game, one player decides how much of an endowment to transfer to another, with the receiving player having an empty choice set. Most players make positive transfers, which Andreoni and Miller (2002) show to be consistent with altruistic preferences.

My results contrast starkly with those of Ayres, Vars, and Zakariya (2005), who show that tips received by black drivers in New Haven, Connecticut are around one-third less than those received by white drivers. However, the study is based on a very small sample (just ten drivers) in a location that might not reflect all markets.[3] In contrast, my dataset covers around 130,000,000 rides from around 30,000 drivers.

Another advantage of my data over that used by Ayres, Vars, and Zakariya (2005) is that it is collected over a long time period. This allows me to examine how tipping behaviour reacts to external events. Namely, I examine whether the white/non-white tipping differential is affected by the April 2013 bombing of the Boston Marathon, an event which I show reinforced negative stereotypes about out-groups. Despite this, I find no change in change in consumers' tipping behaviour towards particular groups around the terror incident. This contrasts with evidence about the effect of terror on attitudes (e.g. Schüller, 2016), and suggests that stated preferences might be more malleable than revealed preferences.

The paper proceeds as follows: in Section 3.2, I introduce the institutional background and data, whilst Section 3.3 outlines the procedure for predicting drivers' race/ethnicity. In Section 3.4 I present a simple model of passenger behaviour where tipping is motivated by altruism and coherence to norms. In Section 3.5 I estimate the differences in tip amounts according to drivers' racial category. In Section 3.6 I detail a number of quality of service measures and show that these do not account for between-race differences in tip amounts. Section 3.7 discuses the (non-)effect of the Boston marathon bombing and Section 3.8 concludes.

3.2 Background and Data

My study is focused on rides made in 'yellow' New York City taxis. These serve the five New York boroughs of Manhattan, Brooklyn, Queens, The Bronx and Staten Island, providing over 450,000 trips per day. Since 2008, these cabs have contained electronic loggers which wirelessly transmit a record of each journey made to the New York Taxi and Livery Company (T&LC), the organisation that licenses vehicles and drivers. Following freedom of information requests made to the T&LC, the data recorded by the journey loggers in all vehicles are in the public domain. Each entry in the dataset corresponds to one journey, and includes a unique identifier for the driver and for the vehicle.

3. Indeed, after clustering standard errors on the driver level, Ayres, Vars, and Zakariya (2005) find that the white/non-white differential is no longer statistically significant.

3.2.1 Ride data

The recorded information for each journey includes the fare (broken down into its component parts of fare, surcharge, tax and tolls), the payment method (cash or credit card), GPS coordinates of the pick-up and drop-off locations, the start and end time of the trip, the distance driven according to the vehicle's odometer, and driver and vehicle identifiers. If the rider chooses to pay by credit card rather than cash (55% of journeys in 2013), the size of the tip is also recorded.

I work with data from 2009 and 2013, where the data disclosures from the T&LC make it possible to match the driver identifier to the T&LC's public data on registered drivers. After excluding observations that fail consistency checks, I match pickup and dropoff coordinates as recorded by taxis' GPS devices to census tracts and neighbourhoods using PostgreSQL.[4] I exclude journeys paid for by cash, as tips are not recorded for these fares. The final dataset includes 125,807,276 rides. Appendix Section C.1 describes the procedure for constructing the final dataset in more detail.

The median fare amount is $10.50. 93% of rides in my dataset originate in Manhattan and 4% at one of New York's airports. Dropoff destinations are somewhat more diffuse, but are also highly concentrated in Manhattan (Appendix Figure C1 and Appendix Figure C2 show maps of pickup and dropoff locations).

3.2.2 Fare and Tip Payment

At the end of a journey, a passenger can pay his or her fare [5] by cash or by using an electronic display and credit card reader installed in the back of the cab. The software used to control the electronic display is provided by one of two vendors.[6] Both vendors offer three suggested tips, and the option of manually entering a dollar amount. In 2013, the three default tip amounts were 20%, 25% and 30% of the fare amount, whilst in 2009 the default tip amounts varied by vendor.[7] Haggag and Paci (2014) show that these technological norms have strong effects on behaviour; indeed, in my dataset, around 60% of tips are at one of the default amounts. Figure 3.1 shows the distribution of tips as a percentage of fare amount in 2013.

4. Neighbourhoods refer to Neighbourhood Tabulation Areas, as defined by the NYC Department of Planning. These areas are aggregations of census tracts into 'neighbourhood-like areas'.

5. The fare charged to a passenger consists of up to four parts. The base is calculated on the basis of distance and time, apart from journeys to/from LaGuardia and JFK airports, which are subject to fixed fares. A surcharge of $1 is applied on journeys taking place in peak weekday hours and of 50 cents on journeys taking place at night. Any bridge or tunnel tolls are passed on to the passenger. Finally, since 2013, a $0.50 'MTA tax' is applied to all journeys ending in New York City or Nassau, Suffolk, Westchester, Rockland, Dutchess, Orange or Putnam Counties.

6. Appendix Figure C3 shows the display of one of the vendors.

7. The software from one vendor displays the default tip amounts only as percentages, the other also as dollar values. In all regressions, I include a fixed effect for the vendor of the payment system.

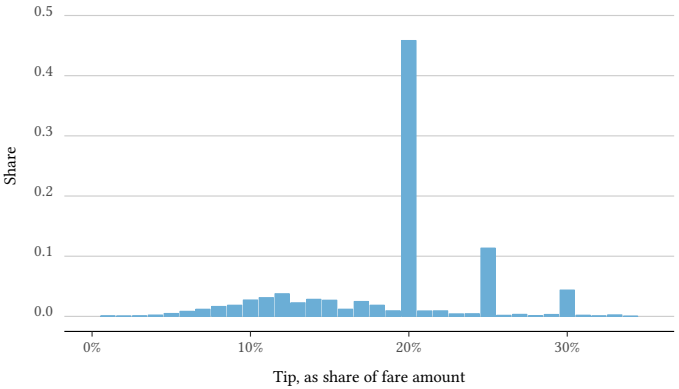


Figure 3.1. This figure shows the empirical distribution of tips, expressed as a share of the fare, in 2013. 'Default' tip amounts suggested by the payment software are 20%, 25% and 30% of the fare.

3.2.3 Drivers

Yellow cab drivers are required to hold a licence from the T&LC. In order to receive a licence, applicants must pass a defensive driving and a driver education course, covering city geography, rules of the T&LC and safe driving. At the time the data I use was collected, drivers also were required to have taken and passed an English proficiency test.

According to aggregate statistics, drivers of yellow cabs are overwhelmingly migrants, with only around 5% of registered drivers being born in the United States as of 2016 (New York Taxi and Livery Company, 2016). The most common countries of birth are Bangladesh, Pakistan, India and Haiti. Drivers are almost exclusively male.

Whilst driving a cab drivers display a license which contains a photograph and their name—this license must be visible to passengers (New York Taxi and Livery Company, 2011, Chapter 80). At least some passengers seem to identify the identity of their driver: speaking to the New York Times following the September 2001 attack in New York, a Muslim driver claimed that covering up his license information led to higher tips (Kennedy, September 24, 2001). Asian drivers have also been victims of hate crimes: in 2013, a white supremacist was gaoled for nine and a half years for stabbing a driver after ascertaining that he was a Muslim (Buettner, June 25, 2013). Asian drivers have also reported being spat upon, cursed at and physically abused by passengers

(Ghosh, June 22, 2012).

A public register of licensed drivers is maintained by the T&LC. This register includes a unique license number, which can be matched to the data on rides, and driver name. Unfortunately, the register does not include race, ethnicity or country of birth—instead, I predict these using driver surname in a procedure described in Section 3.3.

3.2.4 Passengers

There is relatively little published data about passengers of yellow cabs in New York. As of 1993, 73% of trips were made by residents of Manhattan (Schaller, 1993). The remainder of trips were made by residents of New York outer-boroughs (10%), New York suburbs (5%), other places in the U.S. (5%) and other countries (5%). As of 2014, 60% of passengers had incomes of over $50,000 and 42% of $100,000 or more, and passengers were 49% male (New York Taxi and Livery Company, 2014). A report from 1982 states that the median passenger is white (Haberman, March 23, 1982), but no other information is available about passenger race or ethnicity.

No data is available about passengers' views on race, but these can be at least imprecisely proxied by the views of wider populations. The American National Election Studies series, nationally representative surveys of the U.S. population, regularly asks its subjects how warmly they feel towards certain groups on a scale from 0 (very cold and unfavourable) to 100 (very warm and favourable). The data suggests considerable in group favouritism: of the 266 residents of New York State questioned in the 2012 round and answering all relevant questions, 38% report feeling more warmly towards whites than blacks, and 59% report feeling more warmly towards Christians than Muslims.[8] 80% of Manhattan residents taking an online Implicit Association Test hosted by Project Implicit, exhibit implicit anti-black bias.[9].

3.3 Predicting Driver Demographics

3.3.1 Race

Race is widely perceived, although racial categories do not have a basis in genetic differences (Cosmides, Tooby, and Kurzban, 2003). In the U.S. there are currently

8. State of residence is the finest level of geographic detail available in the ANES public use dataset. Appendix Figure C6 shows mean thermology scores towards various groups for New York State residents.

9. Project Implicit carries out research on implicit social cognition. The project operates a website where individuals can test their implicit attitudes on various topics, data from which is then collected and made available to researchers (Xu, Nosek, and Greenwald, 2014) Appendix Figure C4 shows the distribution of 'D-scores' of Manhattan residents in the dataset. D-scores above zero indicates implicit anti-black bias. Those taking the IAT test on the Project Implicit website are not necessarily a representative sample of the Manhattan population as a whole.

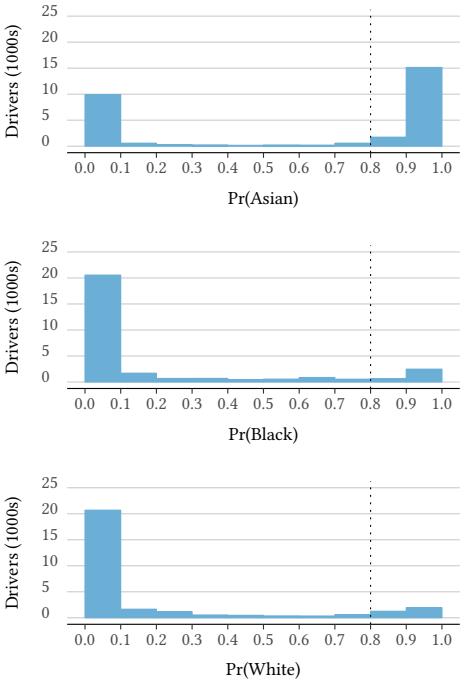


Figure 3.2. Histograms show the distributions of probabilities of membership of three racial categories for drivers in the dataset.

five categories of race defined by the Office of Management and Budget (OMB)[10] : white, black or african american, american indian and Alaskan native, asian, native Hawaiian and other Pacific islander. I do not observe directly in my data which race a driver identifies with (or, indeed, is identified with). Instead, I predict this based on the driver's surname and data from a genealogical website about the global distribution of surnames, aggregate statistics from the T&LC about the birthplaces of drivers, aggregate data derived from the U.S. census about race by country of birth for those resident in the U.S., and aggregate data derived from the U.S. census about self-reported race/ethnicity by surname.

Concretely, let S_i represent the surname of cab driver i, C_i the country of birth of that driver and R_i his racial category, one of *White*, *Black*, *Asian*, *American Indian*, *Native Hawaiian and Other Pacific Islander*. I stands for the set of drivers, S stands for the set of surnames of drivers and C the set of potential countries of birth of those drivers. The procedure aims to estimate $Pr(R_i = r|S_i = s)$ for all $i \in I$, $r \in R$ and $s \in S$.[11]

Firstly, using Bayes' rule, I calculate the posterior probabilities of a driver with a given surname being born in each one of the world's countries, i.e. $Pr(C_i = c|S_i = s$ for each country $c \in C$ and surname $s \in S$,

$$Pr(C_i = c|S_i = s) = \frac{Pr(S_i = s|C_i = c)Pr(C_i = c)}{\sum_{c' \in C} Pr(S_i = s|C_i = c)Pr(C_i = c')} \tag{3.1}$$

Prior probabilities on country of birth, $Pr(C_i = c)$, come from aggregate statistics provided by the T&LC New York Taxi and Livery Company (2014).[12] The probabilities of holding a given surname conditional on a given country of birth, $Pr(S_i = s|C_i = c)$ come from a genealogical website.[13] I make the assumption that there is no selection

10. The OMB, part of the Executive Office of the President of the United States, requires these minimal categories to be used when reporting federal data on race and ethnicity, and are thus used by the U.S. Census Bureau in census questionnaires and reports.

11. U.S. Census data provides $Pr(R_i = r|S_i = s)$, i.e. the share of individuals with a given surname in each racial category, for the population as a whole. However, the population is unrepresentative of New York taxi drivers, who are largely migrants. For example, in the U.S. population most of those with the surname Henri are white, but the procedure I outline here shows that after conditioning on being a New York taxi driver it is most likely that they are black (from Haiti).

12. The data lists the number of licensed drivers born in each of the worlds' countries, excluding, for privacy reasons, countries where less than ten drivers were born. I assume that there are five drivers born in each of these countries. In addition, the list includes an implausibly large 'other' category of over 3,000 drivers, and includes 'Africa-Unspecified' and 'Ussr - Unspecified' as listed countries. I ignore these entries when calculating the share of drivers with any given country of birth.

13. `http://Forebears.io`. The website claims to contain information about the distribution of 12,000,000 names, based on a variety of offline and online genealogical sources. Unfortunately, these sources are not listed. I exclude any surname which, according to the site, exists less than 2,000 times worldwide.

of particular surnames into migration to the U.S. and working as a yellow cab driver in New York.

Next, separately, I estimate the probability of belonging to each racial category, conditional on country of birth and a given surname, $Pr(R_i = r|S_i = s, C_i = c')$. For $c' \neq$ U.S., i.e. birthplaces outside the U.S., I use the 5% sample of the 2000 U.S census (Ruggles et al., 2015). This includes self-reported racial status and country of birth for each sampled individual, and allows me to estimate the proportion of U.S residents from each non-U.S country of birth belonging to each racial category. For $c' =$ U.S, I use a data release from the U.S census bureau which provides the share of individuals with a given surname belonging to each racial category (Comenetz, 2016)[14]:

$$\widehat{Pr(R_i = r|C_i = c, S_i = s)} = \begin{cases} \widehat{Pr(R_i = r|C_i = c)} & , c \neq \text{U.S.} \\ \widehat{Pr(R_i = r|S_i = s)} & , c = \text{U.S.} \end{cases} \tag{3.2}$$

Next, I combine the probabilities in Equation 3.1 and Equation 3.2, leading to the following estimate of $Pr(R_i = r|S_i = s)$:

$$\widehat{Pr(R_i = r|S_i = s)} = \sum_{c' \in C} \widehat{Pr(R_i = r|C_i = c', S_i = s)} \times \widehat{Pr(C_i = c'|S_i)}$$

In this way, I am able to calculate the probability of belonging to each racial category for 33,058 drivers: those whose surnames feature in both the genealogical and census datasets. Figure 3.2 shows histograms of the predicted probabilities of membership of the Asian, black and white racial categories. In each case, the distribution is bimodal, with most mass at the bottom and top of the distribution. The ten surnames most predictive of each racial category are shown in Appendix Table C1.

Finally, I make a prediction of race for each driver, by applying a cut-off at 80%; that is, I assume a driver belongs to a given group if his probability of doing so exceeds 80%. 3,943 drivers are thus classified as white, and 25,967 as non-white (3,386 black and 22,581 Asian). This is consistent with Bangladesh and Pakistan being the most common birth countries of drivers (New York Taxi and Livery Company, 2014).

3.3.2 Ethnicities

In my main results, I compare tips received by drivers classified as white to those classified as black or Asian by the procedure above. This categorisation is based

14. For privacy reasons, only surnames which occur at least 100 times in the U.S. are included in the dataset. Also for privacy reasons, some fields are suppressed for some surnames. When this is the case, I impute the missing fields by reassigning mass evenly to the missing fields. Further, the data release includes Hispanic, an ethnicity, as a separate category. So as not to conflate ethnicity and race, I add the Hispanic share to the white share before proceeding.

on the U.S. census definition of white as 'any person having origins in any of the original peoples of Europe, the Middle East, or North Africa'. It is not a priori clear that this is an appropriate category when trying to identify discriminatory outcomes. Arguing against the use of the term to describe non-minority populations, Bhopal and Donaldson (1998) describe the discord between the U.S. census definition of white and its common use as referring to 'people of European origin with pale complexions'. Testing by the U.S. Census Bureau has shown that many individuals of Hispanic, Arab or Middle Eastern descent do not identify with the white racial category (Mathews et al., 2017), and such individuals have been shown to be subject to discrimination in other settings (e.g. McConnell and Rasul, 2017; Widner and Chicoine, 2011). However, an examination of the surnames assigned to the White category suggest that the category is dominated by drivers of Hispanic, Arab or Middle Eastern descent (see Appendix Table C2).

I therefore further divide the white category, creating two new mutually exclusive subgroups, 'non-Hispanic white' and 'European white'. To do this, I first calculate the probability that a driver with a given surname is of Hispanic ethnicity. This is done using an analogous procedure to that described in Section 3.3.1, but using data on from the U.S. Census Bureau on *ethnicity* by country of birth and surname rather than race in the final step. I also calculate the probability that drivers are of 'European' origin: that is, have birth places in Europe or in the United States. The surnames most predictive of these categories are shown in Appendix Table C3. Finally, I classify drivers as 'non-Hispanic white' if the probability of being both white and non-Hispanic exceeds 80% (2,327 drivers) and as 'European white' if the probability of being white, non-Hispanic and of European-origin exceeds 80% (678 drivers). These new groups will allow me to compare the tips received by white and non-white drivers with less inclusive definitions of the white category.

3.4 Motives for Tipping

In this section I introduce a simple model of tipping. Any such model should account at least for the two main regularities in the data: that the overwhelming majority of rides receive a positive tip, and that there is considerable bunching around the 'default' options provided by the payment systems.

If a customer and a service provider interact repeatedly, customer tipping can be motivated within a classical microeconomic framework as sustaining a high quality of service, positive tipping equilibrium. However, since the probability of repeat interaction between a given driver and passenger is low[15] a more realistic modelling

15. Passengers and drivers coordinate taxi rides by means of street hails: yellow cabs are prohibited from carrying radios which might otherwise be used to coordinate rides. There are 40,000 active drivers in the city. Manhattan's adult residents, of whom two thirds ride taxis at least some of the time, make an

assumption in this setting is that of a single-shot interaction between passenger and driver. This makes tipping difficult to reconcile with narrow self interest. Other potential motivations for tipping are altruism (Saunders and Lynn, 2010; Greenberg, 2014) and adherence to social norms (Azar, 2004). Here I set out a very simple model of tipping which incorporates both of these motives. The model is an adaptation of that of DellaVigna, List, and Malmendier (2012), who model charitable giving under social pressure.

Passenger i receives the following utility from tipping a driver d the amount t:

$$U_i(t) = u(W_i - t) + \alpha_{id} \cdot v(\theta W_d + t) - s(t)$$

$u(W_i - t)$ represents the utility of private consumption given wealth W_i, and is assumed to be satisfy the standard properties of $u'(.) > 0$ and $u''(.) < 0$. The term $\alpha_{id} \cdot v(\theta W_d + t)$ captures the passenger's utility from other-regarding preferences. The α_{id} term reflects the level of altruism of passenger i towards driver d, who has initial wealth W_d. α_{id} could be negative, indicating that the passenger has disutility over the drivers wealth (i.e. malice towards the driver) or positive, the standard case of altruism. The specification of $v(\cdot)$ allows for two different motives for altruism: when $\theta = 1$ the passenger has utility over the whole of the driver's consumption, pure altruism. At the other extreme, $\theta = 0$, the passenger's utility from transferring tip t to the driver comes purely from a 'warm-glow' (Andreoni, 1990). The final term, $s(t)$, captures a cost of tipping less than the norm, g_s. This cost is assumed to be highest when tipping zero, and decreases until the norm is reached. The norm might be social or technological, such as those provided by the default tip options, and is assumed to be independent of the identity of the driver.[16]

DellaVigna, List, and Malmendier (2012) prove that there is a unique optimal tip for each type (i and d) which weakly increases in the level of altruism felt by the passenger to the driver, α_{id}. This means that differences in the tip amounts received by different drivers reflect variations in the level of altruism felt towards them. This shows that tips are informative about passengers level of altruism towards drivers, and by extension, might reveal racial preferences.

Furthermore, three regimes can be characterised, based on the values of α_{id} and the cost of deviating from a norm:

- 'Scrooge tipping': If α_{id} is low enough relative to $s(t)$ passengers tip either zero or a positive amount less than g_s

average of 100 taxi rides per year (Schaller, 2006).

16. No tipping guide suggests that the size of tip should vary according to characteristics of the recipient, and the technological norms provided by the payment software are not dependent on the driver.

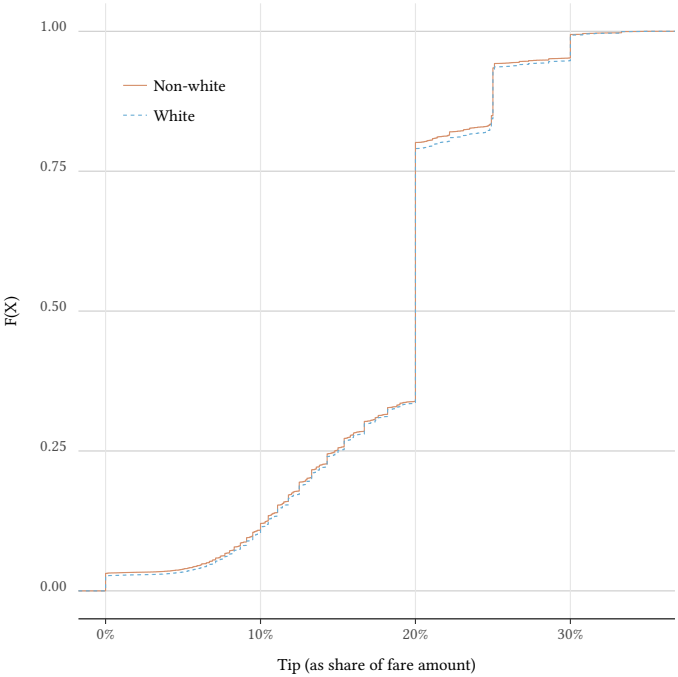


Figure 3.3. Figure shows the empirical cumulative distribution of tip percentages in the year 2013, separately for white and non-white drivers. The default tip options provided by the in-vehicle credit card payment system are 20, 25 and 30% of the fare amount.

- 'Default tipping': The costs from deviating from g are high enough relative to α_{id} to induce tipping at the socially desired level, leading to bunching at g_s
- 'Generous tipping': If the altruism parameter α_{id} is high enough, tips will be higher than g_s.

I will use these three regimes to guide the empirical analysis in the next sections.

3.5 Differentials in Tips

This section provides descriptive evidence on the differential in tips received by white and non-white drivers. As per Section 3.3, I classify drivers as white if the predicted probability that they are white exceeds 80%. I classify them as non-white if the predicted probability of the belonging to the Asian or black categories exceeds the same amount. I limit my data to that of year the 2013 in order to improve computational tractability and because in this year the default tip amounts offered by the credit card payment system were constant across all vehicles.

The empirical distribution of tips received (as a percentage of the fare amount) by white and non-white drivers is shown in Figure 3.3. The overall distributions of tips are very similar, with a maximum difference in the empirical distribution function of 0.01[17]. Noticeably, white drivers appear marginally more likely to receive tips above the lowest technological default, 20%.

In Panel A of Table 3.1 I show the mean tip amount (as a percentage of the fare) and the probabilities of receiving a 'scrooge' (less than 20%, the first default amount), 'default' (20%) or 'generous' (a tip over 20%) tip separately for white and non-white drivers. This panel reveals that white drivers receive a mean tip percentage of 18.5%, compared to 18.7% for non-white drivers, a difference of 0.23 percentage points. On a fare of $10.5, the median in my dataset, this amounts to a two cent difference, corresponding to a yearly income differential of around $100 for a full time driver. Non-white drivers receive 'generous' tips from 20.6% of journeys, compared to 22.2% for white drivers—that is, they are 7% less likely to receive such a tip.

Although these differentials are consistent with discrimination against non-whites, they could be also driven by two types of selection. Firstly, there might be selection of drivers into different types of rides—at different times of day, days of week, or in different areas—which might have different types of passengers or otherwise be tipped differently. In order to control for such potential cofounders, I extract a sample of journeys which are exactly matched on a wide set of journey characteristics. Namely, I attempt to match every ride with a white driver to a ride with a non-white driver

17. Nonetheless, a Kolgmogorv-Smirnov test rejects the hypothesis that the empirical distributions are drawn from the same probability distribution at all conventional levels of significance.

Table 3.1: Summary differences

	White	Non-white	Difference
Panel A: All journeys			
Mean tip (as % of fare)	18.69	18.46	−0.23
Scrooge tip share	0.34	0.34	0.00
Default tip share	0.45	0.46	0.01
Generous tip share	0.22	0.21	−0.01
Observations	7,213,298	48,810,201	
Panel B: Exact matches			
Mean tip (as % of fare)	18.74	18.48	−0.26
Scrooge tip share	0.33	0.34	0.01
Default tip share	0.45	0.46	0.01
Generous tip share	0.22	0.20	−0.02
Observations share	6,828,338	6,828,338	
Panel C: La Guardia pickups			
Mean tip (as % of fare)	19.58	19.47	−0.11
Scrooge tip share	0.32	0.31	−0.01
Default tip share	0.38	0.39	0.02
Generous tip share	0.30	0.30	−0.00
Observations share	171,800	1,383,677	

Notes: The table shows differences in the characteristics of tips received by white/non-white drivers. The figures in Panel A are based on the whole sample of journeys. The figures in Panel B are based on a smaller sample of journeys which are exactly matched on the weekday-hour of the start of journey, the pickup and dropoff neighbourhood pair, the vendor of the vehicle's payment system and fare amount (rounded to the nearest dollar). The figures in Panel C are based on journeys originating from La Guardia airport, where drivers and passengers both queue at taxi stands, resulting in quasi-random matches between drivers and passengers. Scrooge tips are defined as those less than 20%, default tips those of 20% and generous tips those above 20%.

which is identical with regards to: the vendor of the credit card payment system, the weekday-hour of the journey start (e.g. Wednesday, 10am), the pickup and drop-off neighbourhood and the fare amount (rounded to the nearest dollar). I match without replacement and if there are many potential matches, the match is random. In the end, I am left with a sample of 6,828,338 rides; that is, I am able to find a 'non-white' match for the vast majority of journeys with a white driver (7,213,298). Mean comparisons on this matched sample are shown in Panel B of Table 3.1. The difference in average tip received is 0.26 percentage points, with minority drivers around 9% less likely to receive a generous tip, only marginally different to the unmatched sample.[18] As such, the differential is robust to controlling for selection of drivers into different types of journeys.

As well as selection of drivers into certain types of journeys, there might be non-random selection of drivers to riders. There is anecdotal evidence of some drivers refusing to respond to street hails from minority passengers, who have in turn been shown to provide lower tips on average (Lynn et al., 2008). If white drivers were more or less likely to engage in such behaviour, the estimates so far might reflect differential selection of passengers. In order to control for this type of selection, I repeat the analysis, but now limit the analysis to trips originating from LaGuardia airport. Here, both customers and taxis queue in lines, resulting in quasi random assignment of passengers to taxi drivers. Results are shown in Panel C of Table 3.1. Again, non-white drivers are less likely to receive generous tips, but receive an mean tip which, as a share of the fare, is just 0.11 percentage points lower than that received by white drivers.

The evidence so far is for a rather small white/non-white differential in tips received. In order to quantify the magnitude of the differential in relation to overall between driver-variation I regress my outcomes of interest on journey-level controls (an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount) as well as individual driver fixed effects.[19] Formally, I estimate the following regression equation:

$$TipPercentage_i = \alpha_0 + \boldsymbol{\delta} \boldsymbol{X}_j + \boldsymbol{\gamma}_i + u_i$$

where $\boldsymbol{X}_j$ is the vector of journey levels controls, $\boldsymbol{\gamma}_j$ a vector of driver fixed effects and $TipPercentage_i$ is the tip amount as a share of the fare amount. The distribution of $\hat{\boldsymbol{\gamma}}$ is

18. An alternative approach to exact matching is Ordinary Least Squares regression, including journey characteristics as control variables. The results of these regressions, which entail parametric assumptions not required for the matching approach, are presented in Appendix Table C4 and Appendix Table C5. With a full set of journey controls, these show very similar 'treatment effects' to the matching estimators.

19. I limit my sample to rides from drivers with at least 2,000 journeys in the dataset in order to improve the precision of fixed effects estimation.

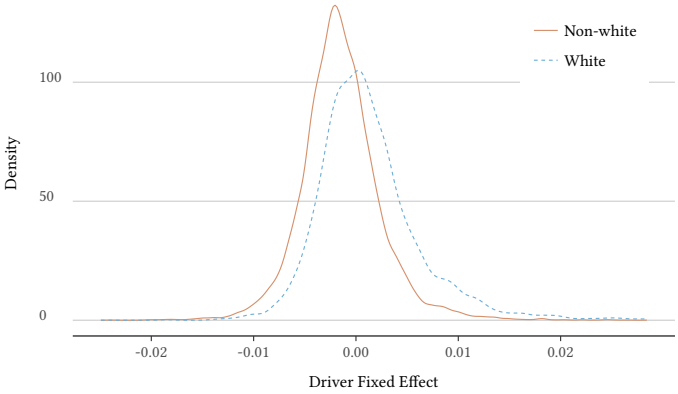


Figure 3.4. Figure shows the distribution of estimated driver fixed effects in a regression on tip amount, as a percentage of fare amount, separately for white and non-white drivers. The fixed effects are estimated from a regression of tip (as a share of fare amount) on journey characteristics and a fixed-effect for each driver with with at least 2,000 journeys in the dataset. The journey-level controls included in the regression are an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount.

plotted in Figure 3.4 separately for white and non-white drivers (rather than omitting the fixed effect for an arbitrary driver, the model is identified by the assumption that the fixed effects sum to zero). The amount of between-driver variation is relatively low, and the distribution of fixed effects for non-white drivers is shifted downwards by around a quarter of a standard deviation compared to white drivers. Race therefore accounts for a relatively small amount of overall between-driver variation.

In Table 3.2, I investigate whether the differentials increase in magnitude when making restrictions on the drivers considered as white. As detailed in Section 3.3.2, the procedure to predict race from surname identifies drivers of Hispanic ethnicity and with birthplaces in the Middle East as white. However, passengers might not identify these passengers as white, or as sharing their identity. In the first three columns of Table 3.2, I exclude any driver who is Hispanic with a probability of more than 20% from the white category. In the second set of columns, I limit the white group to those who, with probability of at least 80%, are born in European countries or the United States. Panels A, B and C show mean differences in the full sample, in exactly matched

Table 3.2: Summary differences: varying definition of white

	Exclude Hispanics			Exclude non-European whites		
	Non-Hispanic white	Non-white	Δ	European white	Non-white	Δ
Panel A: All journeys						
Tip, as % of fare	18.71	18.46	−0.25	18.87	18.46	−0.40
Scrooge tip	0.33	0.34	0.00	0.33	0.34	0.01
Default tip	0.45	0.46	0.01	0.44	0.46	0.01
Generous tip	0.22	0.21	−0.01	0.23	0.21	−0.02
Observations	4,220,165	48,810,201		1,101,637	48,810,201	
Panel B: Exact matches						
Tip, as % of fare	18.77	18.48	−0.29	18.92	18.46	−0.45
Scrooge tip	0.33	0.34	0.01	0.33	0.34	0.02
Default tip	0.45	0.46	0.01	0.45	0.46	0.01
Generous tip	0.22	0.20	−0.02	0.23	0.20	−0.03
Observations	4,005,631	4,005,631		1,044,424	1,044,424	
Panel C: La Guardia pickups						
Tip, as % of fare	19.73	19.47	−0.27	19.95	19.47	−0.48
Scrooge tip	0.31	0.31	0.00	0.31	0.31	−0.00
Default tip	0.38	0.39	0.02	0.36	0.39	0.03
Generous tip	0.31	0.30	−0.02	0.33	0.30	−0.03
Observations	82,684	1,383,677		25,719	1,383,677	

Notes: The table shows differences in the characteristics of tips received by non-Hispanic white/non-white drivers (first three columns) and between European white/non-white drivers (next three columns). The figures in Panel A are based on the whole sample of journeys. The figures in Panel B are based on a smaller sample of journeys which are exactly matched on the weekday-hour of the start of journey, the pickup and dropoff neighbourhood pair, the vendor of the vehicle's payment system and the fare amount (rounded to the nearest dollar). The figures in Panel C are based on journeys originating from La Guardia airport, where drivers and passengers both queue at taxi stands, resulting in quasi-random matches between drivers and passengers. Scrooge tips are defined as those less than 20%, default tips those of 20% and generous tips those above 20%.

journeys and in pickups from La Guardia airport respectively, as per Table 3.1.

Excluding Hispanic whites from the sample has little effect on the white/non-white differential; the difference in average tip as a percentage of fare amount remains close to 0.2 percentage points, as per Table 3.1. However, the differentials when comparing European whites to non-whites—shown in Column 6—are around double the size of those shown in Table 3.1. Nonetheless, they are still small in both absolute and relative terms. As such, even when choosing the definition of white that might a priori be expected to maximise the white/non-white tip differential, I find no evidence for widespread discrimination.

3.5.1 Effects of Defaults

In this subsection I investigate to what extent the default tipping options provided by the payment software influence the white/non-white tipping differential. I do so to investigate whether the presence of default tipping options constrains choice strongly enough to prevent revelation of anti-minority preferences.

To do so, I use the 2009 ride level data from the T&LC. During 2009, the credit card payment system provided by VTS, one of the two vendors of credit card payment system, offered different default tip amounts depending on the size of the fare. For fare amounts less than \$15, default tip options of \$2, \$3 and \$4 were provided. For fare amounts above \$15, defaults were calculated as 20%, 25% and 30% of the fare. As shown in Figure 3.5, this leads to a discontinuity in the share of rides tipped at one of the default options, with around 75% of rides tipped at a default amount before the discontinuity and only 25% after. In Table 3.3, I exploit this discontinuity to test for the effects of defaults on the white/non-white differential.[20] I define an indicator variable that takes the value of one for fares less than \$15, and interact this with a non-white indicator variable indicating whether a given driver is non-white, leading to the following regression specification:

$$\begin{aligned} TipPercentage_i = \beta_0 + \beta_1 NonWhite_j + \beta_2 \cdot \mathbb{1}[Fare_i \leq \$15] \\ + \beta_3 NonWhite_j \cdot \mathbb{1}[Fare_i \leq \$15] + \boldsymbol{\delta X}_i + u_{ij} \end{aligned} \tag{3.3}$$

where i indexes a ride provided by driver j, $NonWhite_j$ is a binary variable indicating whether the driver is non-white, and $\boldsymbol{X}_i$ is a vector of journey levels controls. I limit the sample to rides with fares between \$13 and \$17 – that is, a two dollar window around the discontinuity. Results are presented in Table 3.3. The coefficient on $NonWhite_j \cdot \mathbb{1}[Fare_i \leq \$15]$ is close to zero and far from being statistically different from zero at

20. Admittedly, I can only draw conclusions around behaviour at this single discontinuity, but this provides suggestive evidence of the wider effects of defaults.

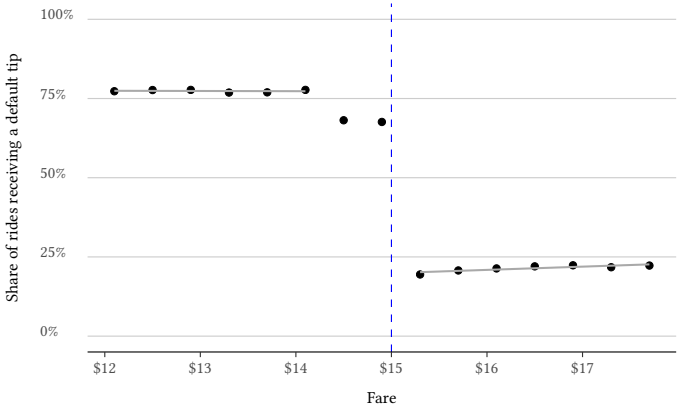


Figure 3.5. Figure shows the share of rides that receive a default tip amount at various fare amounts, for rides taken in 2009 in vehicles equiped with a VTS-supplied payment system. The default tip options on fares less than $15 were $1, $2 and $3 dollars, the default on fares above $15 were 20%, 25% and 30%.

any conventional level of significance. As such, there is no evidence that exogenously changing the share of rides receiving a non-default tip changes the white/non-white tip differential and suggests that the presence of default tipping options is not responsible for the relatively small difference in tips received by white/non-white drivers.

3.6 Quality of Service

In this section I investigate whether the differential described in Section 3.5 reflects differential quality of service across racial groups. In a meta-study, Lynn and McCall (2000) find that tips respond to service quality only very weakly. Nonetheless, I introduce a number of quality of service measures and investigate to what extent the measured tip differential is related to variation in these measures. Three quality of service measures—number of violations of T&LC rules, experience, and willingness to engage in fraud—are measured on the driver level. Two quality of service measures—vehicle characteristics and trip time—are measured on the journey level.

Each column in Table 3.4 shows the impact of controlling for a given quality of service measure on the white/non-white tip differential. Because most controls necessitate limiting the sample in some way, Panel A shows a baseline regression without the quality of service control, whereas the regression reported in Panel B includes it. In

Table 3.3: Effect of defaults

	Dependent variable:	
	(1) Tip as % of fare	(2) Generous tip dummy
Non-white	−0.34***	−0.0184***
	(0.04)	(0.002)
Fare ≤ $15	−0.71***	0.1343***
	(0.04)	(0.003)
Non-white × Fare ≤ $15	−0.014	−0.0033
	(0.04)	(0.003)
Journey-level controls	✓	✓
Observations	947,700	947,700

Notes: Columns reports the estimated coefficients and standard errors from an Ordinary Least Square regressions. Standard errors are clustered at the driver level. Controls are an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

both cases, I include a standard set of journey-level controls: an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount.

3.6.1 Driver Summons

The T&LC issues summons against drivers in response to consumer complaints or violations of T&LC rules uncovered by enforcement officers working in the field. Summons are then heard in court, and fines can be issued to drivers found guilty. Following a freedom of information request, the T&LC provided me information on all recorded driver summons for the years 2010-2016. The three most common reasons for summons are violations of rules on soliciting passengers, allowing a taxi to be used for purposes other than authorised rides, and unjustified refusal to transport passengers. In Column 1 of Table 3.4, I include the number of summons received by a driver in 2013 as an additional control; the median number of summons per driver is one, the highest number of summons for a driver is 23. Because there might be a mechanical connection between number of rides made and number of summons, I also control for the number of rides made by the driver in 2013. The results from this specification are shown in Panel B of Column 1. Drivers who receive more complaints also receive lower tips, which suggests that the control variable indeed measures quality of service. However, comparing the coefficient on non-white between Panel A and Panel B shows that including the control makes no difference to the estimated white/non-white coefficient.

3.6.2 Driver Overcharging

In 2010, it came to light that a subset of drivers had been systematically overcharging riders by illegally engaging an 'out-of-city' rate mode on the taxi meter. This mode, which drivers should only be activated outside New York city limits, doubles the rate at which the fare accrues, resulting in passengers being illegally overcharged for their journeys. I count the number of times each driver defrauded passengers in 2009 this way, which I then take as a proxy for general service orientation.

In order to identify fraudulently charged rides, I limit the 2009 data to those rides starting and ending in the city limits, i.e. rides for which the out-of-city rate code should never be applied. I then calculate the theoretical maximum amount that could have been legitimately have been charged for the journey: the base rate of \$2.50, plus \$0.40 a unit, which accrue every fifth of a mile or every minute travelling at less than 12 miles per hour. If the actual fare is higher than that theoretical maximum, I code the journey as fraudulent:

$$Fraud_i = \mathbb{1}[Fare_i > 2.5 + 0.4 \times (Distance_i/5 + TripTime_i)]$$

where distance is measured in miles and time in minutes. Overall, I identify 9,287 fraudulently charged rides across 5,020 drivers. For each driver, I divide the number of fraudulent rides by the number of rides within the city limits, and include this as the quality-of-service control control. Results are shown in Panel B of Column 2 of Table 3.4. The relationship between fraudulent behaviour and tipping turns out to be rather weak, and the coefficient on non-white is unchanged compared to a baseline specification.

3.6.3 Driver Experience

Haggag, McManus, and Paci (2017) show that experience increases drivers' per shift income, largely by helping them find new passengers quickly after the end of a ride. Experience might also improve the quality of service provided to passengers by, for example, improving knowledge of New York's geography. Unfortunately, I do not observe how long any given driver has been registered with the T&LC or the cumulative number of rides they have given.

Instead, following the approach suggested by Haggag, McManus, and Paci (2017), I limit my sample to rides from drivers who make no rides between 1 January 2013 and 31 March 2013. The assumption is that drivers who enter the dataset after this point have no previous experience as a New York taxi driver. I then generate a variable that, for every ride, measures the cumulative number of rides given by the driver up until the present ride. I include this variable as an additional control. Results of this specification are shown in Panel B of Column 3 of Table 3.4. As might be expected, new drivers receive higher tips as they gain experience, but the control leaves the coefficient on non-white almost unchanged.

3.6.4 Trip Time

Passengers have an interest in their driver taking an optimal route, reducing their time in the taxi and the fare they pay. Drivers, however, might have an interest in taking a longer route, increasing their revenue.[21] For every trip in 2013, I use the 'Open Source Routing Machine' software package to calculate the fastest route between the pick-up and drop-off location, the distance of this route, and the expected duration of the journey.[22] The optimal route is calculated on the basis of OpenStreetMap map data and assumes light traffic conditions. For each journey, I calculate the deviation between log optimal travel time and log actual travel time, and include this in the regression as a quality of service control. In order to control for variation in traffic conditions, I limit my analysis to trips which both start and end within Manhattan

21. Drivers spend a considerable amount of time without passengers, looking for fares.

22. The Open Source Routing Machine is available at `http://project-osrm.org/`. It uses contraction hierarchies to calculate the optimal route through a network; this is considerably more efficient than Dijkstra's algorithm.

Table 3.4: Quality of service

	Dependent variable:				
	Tip, as percentage of fare				
	(1)	(2)	(3)	(4)	(5)
Panel A: Without quality of service control					
Non-white	−0.27***	−0.21***	−0.37***	−0.30***	−0.27***
	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)
Panel B: With quality of service control					
Non-white	−0.27***	−0.21***	−0.38***	−0.28***	−0.29***
	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)
Complaints	−0.22***				
	(0.04)				
Share of journeys fraudulent		−2.16*			
		(1.31)			
Cumulative rides (log)			0.16***		
			(0.01)		
Extra trip time				−1.17***	
				(0.41)	
Vehicle fixed effects					✓
Baseline controls	✓	✓	✓	✓	✓
Controls number of journeys	✓				
Hour of year F.E				✓	
Observations	55,854,577	32,442,933	2,602,449	55,854,577	46,505,989

Notes: Each column in each panel reports the coefficients and standard errors (in brackets) of an Ordinary Least Squares regression. Standard errors are clustered at the driver level. Baseline controls are an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount. The sample for Column 1 and Column 5 is all journeys in 2013. The sample for Column 2 is all journeys by drivers who were active in 2009. The sample for Column 3 is all journeys by drivers who enter the dataset on or after 1 April 2013. The sample for Column 6 is all 2013 journeys which can be matched to a vehicle. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

and include a fixed effect for each of the 8,760 hours in the year. The results from this specification are shown in Panel B of Column 3 of Table 3.4. The estimated coefficient on the control is significant and negative, showing that tips are lower when a journey takes longer than optimal, but again, this does not seem to drive the white/non-white tip differential.

3.6.5 Vehicle Characteristics

Vehicle quality—for instance cleanliness or age—might affect passengers' enjoyment of journeys. If for whatever reason vehicle quality is correlated with drivers' race, then the estimate of the white/non-white tip differential might suffer from missing variable bias. In order to examine if differential vehicle quality drives the differentials shown in Table 3.1, I exploit the fact fact that most vehicles are shared by multiple drivers. The journey-level dataset includes a unique vehicle identifier, which I include as a fixed effect in my regression. As a result, the regression estimation estimates the expected white/non-white differential whilst holding the vehicle constant. These results are shown in Column 5 of Panel B of Table 3.4, with baseline results without vehicle fixed effects shown for comparison in Panel A. The coefficient of the non-white dummy is not changed by the inclusion of vehicle fixed effects.

Admittedly, none of these control variables are able to capture all dimensions of service quality. However, the overall evidence from this section is that the magnitude of the white/non-white tip differential is not a reflection of differential quality of service between drivers of different racial groups.

3.7 Response to Terror

Finally, in this section, I investigate whether the white/non-white differential in tips was impacted by a widely reported terror attack in the nearby state of Massachusetts. Terror attacks have been shown to increase anti-minority sentiment as measured by surveys, with effects even found far away from the locus of the attacks: Åslund and Rooth (2005) and Schüller (2016) show effects in Sweden and Germany, respectively, of the attacks in New York on 11th September 2001. There is also evidence on changes in behaviour caused by terror attacks. Gould and Klor (2016) find an increase in anti-Muslim hate crime across the U.S after the '9/11' attack. Increase in anti-minority sentiment is not necessarily limited to one group: McConnell and Rasul (2017) shows that judges give longer sentences to *Hispanic* defendants after the 11th September 2001 attacks, perpetuated by Arab Muslims.

On the afternoon of 15th April 2013 (Patriots' Day) two bombs were detonated at the finish line of the Boston Marathon, killing three and injuring nearly 300. The events were widely reported throughout the U.S. Many assumed that the bombing

Table 3.5: Effect of Boston marathon attack

	Treatment period:			
	Same day (1)	Three days (2)	One week (3)	Two weeks (4)
Panel B: Dependent variable: tip, as percentage of fare				
Non-white	−0.43***	−0.43***	−0.43***	−0.43***
	(0.04)	(0.04)	(0.04)	(0.04)
Boston	−0.02	−0.14	−0.05	−0.02
	(0.05)	(0.09)	(0.06)	(0.05)
Non-white × Boston	0.0003	0.0012	0.0005	0.0003
	(0.05)	(0.09)	(0.06)	(0.05)
Panel A: Dependent variable: generous tip dummy				
Non-white	−0.0271***	−0.0273***	−0.0272***	−0.0271***
	(0.0041)	(0.0040)	(0.0040)	(0.0041)
Boston	0.0001	−0.0033	−0.0006	0.0001
	(0.0031)	(0.0056)	(0.0040)	(0.0031)
Non-white × Boston	−0.0012	0.0009	−0.0008	−0.0012
	(0.0031)	(0.0056)	(0.0041)	(0.0031)
Journey-level controls	✓	✓	✓	✓
Observations	13,889,552	13,889,552	13,889,552	13,889,552

Notes: Each column in each panel reports the estimated coefficients and standard errors (in brackets) from an Ordinary Least Square regression. Standard errors are clustered at the driver level. Journey-level controls are the an indicator variable for the weekday-hour of the journey time, an indicator variable for the pickup and dropoff neighbourhood pair, an indicator variable for the vendor of the vehicle's payment system and the fare amount. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

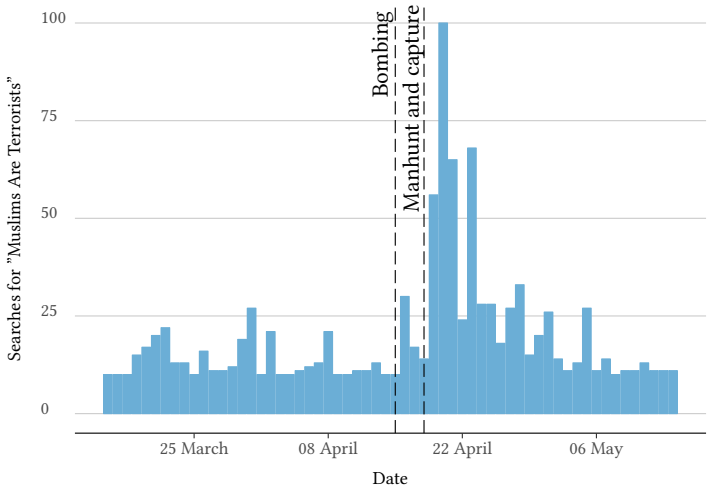


Figure 3.6. Figure shows an index of the quantity of Google searches for the term 'Muslims are Terrorists' around the 15 April 2013 bombing in Boston. Only searches made in New York State are included in the index.

was an act of Islamic terror: a New York Post article wrongly identified the suspect as Saudi National and a Fox News commentator, Erik Rush, posted on Twitter "Yes, they [Muslims] are evil. Let's kill them all." (Hall, April 16, 2013). The identities of the suspects as brothers of Chechen decent was revealed on 18th April 2013, with many news reports focussing on the brothers' Muslim identities (see e.g. Graff, April 19, 2013).

The attack was certainly perceived in New York. Using data from wellness module of the American Time Use Survey, Clark and Stancanelli (2016) show that the Boston bombings caused a large drop in self-reported happiness in states near Boston, including New York. Figure 3.6 shows that the number of Google search terms for 'Muslims are terrorists' increase temporarily in New York state after the bombing.[23] I take this as evidence that the bombings reinforced negative stereotypes against Muslims and,

23. A formal structural break test (Bai and Perron, 2003) identifies breaks in the time series on 15 April 2013, the day of the bombing, and on 29 April 2013, two weeks afterwards.

perhaps, minorities more generally.

In order to investigate whether the attack had effects on tipping behaviour, Table 3.5 reports results from estimating equations of the form:

$$\begin{aligned} TipPercentage_{ij} = \beta_0 &+ \beta_1 NonWhite_j + \beta_2 Boston_i \\ &+ \beta_3 NonWhite_j \cdot Boston_i + \boldsymbol{\delta X}_{ij} + u_{ij} \end{aligned}$$

where $Boston_i$ is a dummy variable indicating whether a ride took place within a given time period of the Boston bombings, and other variables are defined as per Equation 3.3. Because it is not a priori clear how long any treatment effect might last, I vary the definition of $Boston_i$ across the four columns of the table. In Column 1, I consider a ride as 'treated' if it took place on the same day (but after) the attack and in Columns 2 to 4 I include rides in the following three days, one week and two weeks respectively. Across the specifications, there is no sign of increased anti non-white discrimination following the terror attack, with coefficients small and close to zero.

To investigate the effects on a narrower group, I code drivers as of Muslim origin if they have a first name of 'Mohammed', 'Mohammad', 'Muhammad', or 'MD'. I then repeat the empirical exercise, this time identifying how the Muslim/white differential changes in response to the Boston marathon attacks. Again here, there is no statistically significant effect of the terror attack on the tip differential, as reported in Appendix Table C6.

Overall, the evidence is that passengers do not respond to the event in making their tipping decisions. This is consistent with the earlier evidence that drivers' outward identity is not a relevant factor for most passengers when deciding on the size of tip to leave.

3.8 Conclusion

Using a very large dataset, this paper has identified a small differential in the tips received by New York taxi drivers according to their race: non-white drivers receive tips which are, as a share of the fare amount, around 0.2 percentage points lower than white drivers; they lose about $100 a year in income as a result. Most rides, however, are tipped similarly between white and non-white drivers.

I showed that the differential is not significantly changed by controlling for variations in drivers' quality of service, as measured in a variety of ways: drivers' summons and experience, their propensity to commit fraud, the routes they select and the vehicle they drive. By controlling for a number of journey characteristics, I also demonstrated

that the differential does not reflect selection into certain kinds of trips by certain drivers. The differential does not change around a discontinuous break in the share of journeys receiving a 'default' tip.

If quality of service does not vary along unobserved dimensions in ways which correlate with race, then the differential can be seen as reflecting some passengers' favourable treatment of white drivers over non-white. However, the differential is much smaller than that found in the earlier literature, and much smaller than the 'effect' of being a racial minority in Mincerian wage regressions. The paper provides evidence against pervasive taste-based racial discrimination, in this setting at least.

The result is encouraging, although perhaps surprising given earlier work on the topic. What explains the difference between my findings and those of (Ayres, Vars, and Zakariya, 2005)? Here, my data unfortunately do not allow any more than informed speculation. One important factor might be that drivers of New York taxis are predominantly non-white; passengers are likely to have frequent interactions with non-white drivers, perhaps moderating any prejudices, as per Allport's (1954) hypothesis. Alternatively, passengers might not recognise even 'European white' drivers as part of their in-group, these being mostly migrants to the U.S. It is left for further research to understand under which conditions taste-base discrimination occurs.

Appendix C.1: Dataset Construction

The full data disclosure for 2009 and 2013 comprises 344,075,814 journeys. Of these 321,340,791 (93%) can be matched to a driver in the TLC driver database, and 314,967,537 have coordinates for pickup and dropoff locations within the boundaries of New York (and Newark Airport). 129,099,008 of these are fares paid by credit card, where the tip is recorded. Finally, I exclude any rides which fail one or more of the consistency checks suggested by Haggag and Paci (2014):

1. Part of a shift of more than 20 hours (2,177,262 cases)
2. Part of a shift of less than 30 minutes (50,766 cases)
3. Ride duration is zero or longer than 3 hours (168,732 cases)
4. Ride distance is zero or more than 100 miles (530,191 cases)
5. Recorder pickup time is after recorded dropoff time (121,971 cases)
6. Multiple vehicles used within one driver shift (608,177 cases)

This leaves a dataset of 125,807,276 valid journeys.

Appendix C.2: Extra Figures and Tables

Table C1: The ten surnames most predictive of each racial category

	White	Black	Asian
1.	Kowalczuk	Casseus	Shaikat
2.	Yeghiazaryan	Exantus	Newaz
3.	Toader	Estime	Rokonuzzaman
4.	Stoian	Lindor	Foysal
5.	Petrescu	Lafontant	Al-mahmud
6.	Trif	Fils-aime	Ashaduzzaman
7.	Enache	Calixte	Saroar
8.	Baciu	Bazile	Aktaruzzaman
9.	Novac	Bien-aime	Touhid
10.	Grigoras	Cantave	Rasheduzzaman

Table C2: Most likely country of birth, white drivers

Country of birth	Count
Ecuador	397
Dominican Republic	392
United States	278
Colombia	266
Morocco	219
Uzbekistan	179
Turkey	156
Algeria	101
Romania	94
Puerto Rico	83
Russia	79
Tajikistan	38
Bulgaria	37
Poland	32
Afghanistan	29
Peru	27
Ukraine	22
Israel	20
Other	155

Table C3: The ten surnames most predictive of ethnicity, white race

	Hispanic	European
1.	Grisales	Slone
2.	Atehortua	Whalen
3.	Quiceno	Houser
4.	Manjarrez	Huffman
5.	Inoa	Branham
6.	Suriel	Scarlat
7.	Liriano	O'Keefe
8.	Ospina	Niculae
9.	Lizarazo	Stoian
10.	Taveras	Puiu

Table C4: Effect of racial category on tip amount, as percentage of fare

	Dependent variable: tip, as percentage of fare						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Sample: all							
Non-white	−0.23***	−0.21***	−0.29***	−0.21***	−0.20***	−0.21***	−0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	56,021,426	56,021,426	56,021,426	56,021,426	56,021,426	55,854,577	55,854,577
Panel B: Sample: non-whites and non-Hispanic whites							
Non-white	−0.25***	−0.24***	−0.30***	−0.24***	−0.25***	−0.24***	−0.29***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	53,028,456	53,028,456	53,028,456	53,028,456	53,028,456	52,870,388	52,870,388
Panel C: Sample: non-whites and European whites							
Non-white	−0.40***	−0.40***	−0.48***	−0.39***	−0.39***	−0.39***	−0.46***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Observations	49,910,083	49,910,083	49,910,083	49,910,083	49,910,083	49,760,772	49,760,772
Fare amount		✓					✓
Vendor of payment system			✓				✓
Day of year				✓			✓
Weekday hour					✓		✓
Pickup and dropoff neighbourhood pair						✓	✓

Notes: Each column in each panel reports the estimated coefficients and standard errors (in brackets) from an Ordinary Least Square regression. Standard errors are clustered at the driver level. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table C5: Effect of racial category on the probability of receiving a generous tip

	Dependent variable: generous tip dummy						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Sample: all							
Non-white	−0.0120***	−0.0119***	−0.0181***	−0.0113***	−0.0101***	−0.0121***	−0.0168***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0010)
Observations	56,021,426	56,021,426	56,021,426	56,021,426	56,021,426	55,854,577	55,854,577
Panel B: Sample: non-whites and non-Hispanic whites							
Non-white	−0.0134***	−0.0133***	−0.0183***	−0.0131***	−0.0126***	−0.0140***	−0.0184***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0013)
Observations	53,028,456	53,028,456	53,028,456	53,028,456	53,028,456	52,870,388	52,870,388
Panel C: Sample: non-whites and European whites							
Non-white	−0.0232***	−0.0232***	−0.0303***	−0.0227***	−0.0213***	−0.0229***	−0.0290***
	(0.0028)	(0.0028)	(0.0027)	(0.0028)	(0.0028)	(0.0027)	(0.0026)
Observations	49,910,083	49,910,083	49,910,083	49,910,083	49,910,083	49,760,772	49,760,772
Fare amount		✓					✓
Vendor of payment system			✓				✓
Day of year				✓			✓
Weekday hour					✓		✓
Pickup and dropoff neighbourhood pair						✓	✓

Notes: Each column in each panel reports the estimated coefficients and standard errors (in brackets) from an Ordinary Least Square regression. Standard errors are clustered at the driver level. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table C6: Effect of Boston marathon bombing: Muslim-white differential

	Treatment period:			
	(1) Same day	(2) Three days	(3) One week	(4) Two weeks
Panel A: Dependent variable: tip, as percentage of fare				
Muslim	−0.47***	−0.48***	−0.47***	−0.48***
	(0.03)	(0.03)	(0.03)	(0.03)
Boston	−0.26	−0.09	−0.01	−0.01
	(0.16)	(0.06)	(0.04)	(0.03)
Muslim × Boston	0.22	0.08	0.0001	0.03
	(0.17)	(0.06)	(0.04)	(0.04)
Panel B: Dependent variable: generous tip dummy				
Muslim	−0.0329***	−0.0330***	−0.0329***	−0.0331***
	(0.0028)	(0.0028)	(0.0028)	(0.0028)
Boston	−0.0096	−0.0058	−0.0014	−0.0016
	(0.0098)	(0.0039)	(0.0028)	(0.0021)
Mohammed × Boston	0.0066	0.0038	0.0001	0.0016
	(0.0102)	(0.0040)	(0.0029)	(0.0022)
Journey-level controls	✓	✓	✓	✓
Observations	3,940,190	3,940,190	3,940,190	3,940,190

Notes: Each column in each panel reports the estimated coefficients and standard errors from an Ordinary Least Square regression. Standard errors are clustered at the driver level. Journey-level controls include: weekday-hour, software vendor, fare amount, and pickup and dropoff neighbourhood. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

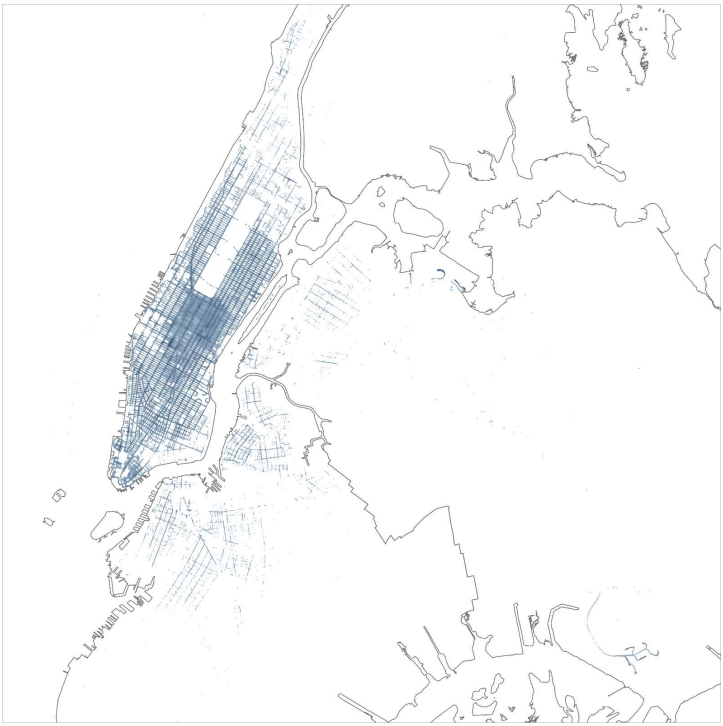


Figure C1. Figure provides a visualisation of the number of pickups in cells of approximately ten by ten meters, with darker colours indicating more pickups. Only cells with at least 100 pickups are included. Most pickups take place within Manhattan. A concentration of pickups at JFK and La Guardia airports is also visible.

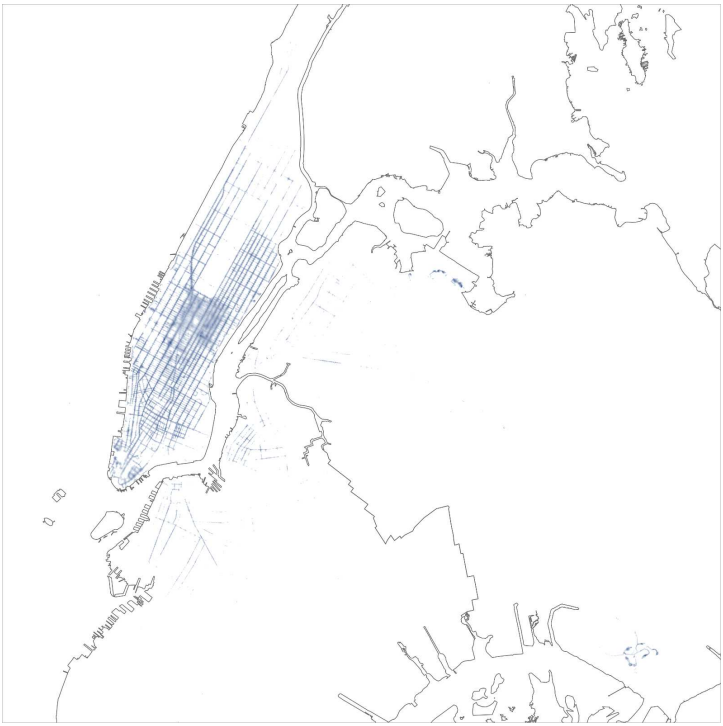


Figure C2. Figure provides a visualisation of the number of dropoffs in cells of approximately ten by ten meters, with darker colours indicating more dropoffs. Only cells with at least 100 dropoffs are included. Most dropoffs take place within Manhattan. A concentration of dropoffs at JFK and La Guardia airports is also visible.

Figure C3. Screen from the credit card payment system of one vendor. Passenger can select a default tip amount (20%, 25% or 30%) or enter a dollar amount. Source: Good Experience, `http://goodexperience.com/blog/2011/02/how-a-taxi-button-cha.php`, last accessed September 2017.

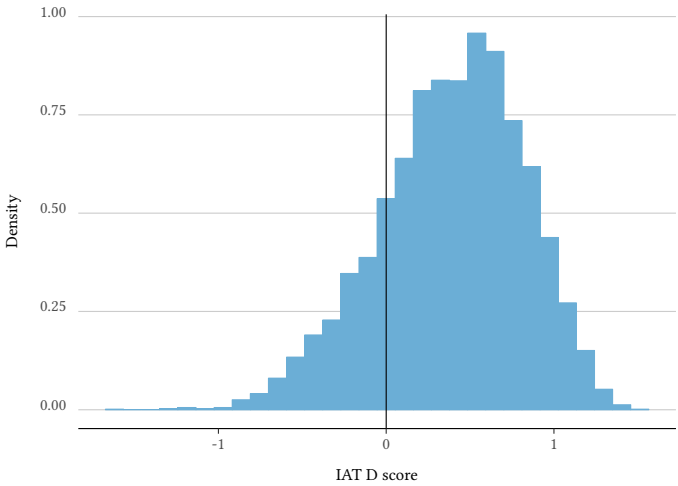


Figure C4. The figure shows IAT scores of Manhattan residents, according to Project Implicit's Race IAT dataset, available via the Open Science Framework at `https://osf.io/y9hiq/`. The figure shows the distribution of IAT 'D-scores' from individuals taking an race IAT test on Project Implicit's website and reporting being resident in New York County (Manhattan). A D-score above zero indicates implicit anti-black bias.

Figure C5. ANES feeling thermometers, New York State residents

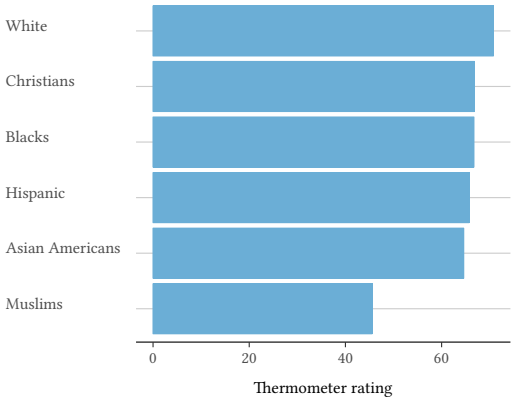


Figure C6. This figure is derived from the 2012 Time Series dataset from the American Nation Election Studies project. The figure shows average scores from feeling thermometers towards given groups for residents of New York state (302). The scale is from 0 ('very cold or unfavourable feeling') to 100 'very warm or favourable feeling'. Missing responses are exluded from the calcluation.

BIBLIOGRAPHY

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "The Political Legacy of American Slavery". *Journal of Politics* 78 (3): 621–641.

Agerström, Jens, and Dan-Olof Rooth. 2011. "The Role of Automatic Obesity Stereotypes in Real Hiring Discrimination". *Journal of Applied Psychology* 96 (4): 790–805.

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30 (2): 175–187.

Alesina, Alberto, and Nicola Fuchs-Schündeln. 2007. "Good Bye Lenin (or Not?): The Effect of Communism on People's Preferences". *American Economic Review* 97 (4): 1507–1528.

Alesina, Alberto, and Paola Giuliano. 2010. "The Power of the Family." *Journal of Economic Growth* 15 (2): 93–125.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough". *Quarterly Journal of Economics* 128 (2): 469–530.

Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review* 100 (5): 2060–92.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, Mass.: Addison-Wesley.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving". *Economic Journal* 100 (401): 464–477.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–753.

Arzheimer, Kai. 2008. *Die Wähler der extremen Rechten 1980 - 2002*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Arzheimer, Kai. 2009. "Contextual Factors and the Extreme Right Vote in Western Europe, 1980–2002." *American Journal of Political Science* 53 (2): 259–275.

Arzheimer, Kai, and Elizabeth Carter. 2006. "Political Opportunity Structures and Right-Wing Extremist Party Success". *European Journal of Political Research* 45:419–446.

Åslund, Olof, and Dan-Olof Rooth. 2005. "Shifts in Attitudes and Labor Market Discrimination: Swedish Experiences after 9-11." *Journal of Population Economics* 18 (4): 603–629.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. 2016. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure". *National Bureau of Economic Research Working Paper Series* 22637 (September).

Avdeenko, Alexandra, and Thomas Siedler. 2017. "Intergenerational Correlations of Extreme Right-Wing Party Preferences and Attitudes toward Immigration". *Scandinavian Journal of Economics* 119 (3): 768–800.

Ayres, Ian, Fredrick E Vars, and Nasser Zakariya. 2005. "To Insure Prejudice: Racial Disparities in Taxicab Tipping". *Yale Law Journal:* 1613–1674.

Azar, Ofer H. 2004. "What Sustains Social Norms and How They Evolve?: The Case of Tipping". *Journal of Economic Behavior & Organization* 54 (1): 49–64.

Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18 (1): 1–22.

BBC. 2001. *BNP: Under the Skin*. `http://news.bbc.co.uk/hi/english/static/in_depth/programmes/2001/bnp_special/roots/1984.stm`.

———. 2005. *Are You Racist? The Test That Claims to Know*. `http://news.bbc.co.uk/2/hi/uk_news/magazine/4447471.stm`.

Beck, Paul Allen, and M. Kent Jennings. 1991. "Family Traditions, Political Periods, and the Development of Partisan Orientations". *Journal of Politics* 53 (3): 742–763.

Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press.

Becker, Sascha O., Katrin Boeckh, Christa Hainz, and Ludger Woessmann. 2016. "The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy." *Economic Journal* 126 (590): 40–74.

Becker, Sascha O., and Luigi Pascali. 2016. "Religion, Division of Labor and Conflict: Anti-Semitism in German Regions over 600 Years". *University of Warwick Working Paper Series* 288 (April).

Bertrand, Marianne, and Esther Duflo. 2017. "Field Experiments on Discrimination." In *Handbook of Field Experiments.*

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94, no. 4 (September): 991–1013.

Bhopal, Raj, and Liam Donaldson. 1998. "White, European, Western, Caucasian, or What? Inappropriate Labeling in Research on Race, Ethnicity, and Health." *American Journal of Public Health* 88 (9): 1303–1307.

Biggs, Michael, and Steven Knauss. 2011. "Explaining Membership in the British National Party: A Multilevel Analysis of Contact and Threat." *European Sociological Review* 28 (5): 633–646.

Bisin, Alberto, and Thierry Verdier. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences". *Journal of Economic Theory* 97 (2): 298–319.

Boisjoly, Johanne, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity". *American Economic Review* 96 (5): 1890–1905.

Buettner, Russ. 2013. "Man Sentenced to 9½ Years in '10 Attack on Cabdriver." *New York Times,* June 25.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83 (401): 123–127.

Carell, S, Mark Hoekstra, and J West. 2016. "The Impact of College Diversity on Behavior Toward Minorities". *NBER working paper* 20940.

Clark, Andrew, and Elena Stancanelli. 2016. "Individual Well-Being and the Allocation of Time Before and After the Boston Marathon Terrorist Bombing." Working paper or preprint, April.

Comenetz, Joshua. 2016. *Frequently Occuring Surnames in the 2010 Census*. United States Census Bureau.

Cosmides, Leda, John Tooby, and Robert Kurzban. 2003. "Perceptions of Race". *Trends in Cognitive Sciences* 7 (4): 173–179.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving". *Quarterly Journal of Economics* 127 (1).

Dippel, Christian, Robert Gold, and Stephan Heblich. 2016. "Globalization and Its (Dis-)Content: Trade Shocks and Voting Behavior". *California Center for Population Research Working Paper Series* 19 (February).

Doleac, Jennifer L, and Luke CD Stein. 2013. "The Visible Hand: Race and Online Market Outcomes". *Economic Journal* 123 (572): F469–F492.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm. 2016. "Refugee Migration and Electoral Outcomes". *CReAM Discussion Paper Series* 19/16.

Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 9, no. 2 (April): 1–22.

Falter, Jürgen W., and Dirk Hänisch. 1990. *Election and Social Data of the Districts and Municipalities of the German Empire from 1920 to 1933.* GESIS Data Archive, Cologne. Data file.

Fernández, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility". *American Economic Journal: Macroeconomics* 1 (1): 146–177.

Fisman, Ray, Yasushi Hamao, and Yongxiang Wang. 2014. "Nationalism and Economic Exchange: Evidence from Shocks to Sino-Japanese Relations". *Review of Financial Studies* 27 (9): 2626–2660.

Fouka, Vasiliki, and Hans-Joachim Voth. 2016. "Reprisals Remembered: German-Greek Conflict and Car Sales during the Euro Crisis". *CEPR Working Paper Series* 9704 (September).

Ghosh, Palash. 2012. "South Asian Taxi Drivers: Victims And Perpetrators Of Racism." *International Business Times,* June 22.

Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5 (5): 927–952.

Giuliano, Paola, and Nathan Nunn. 2017. "Understanding Cultural Persistence and Change". *NBER Working Paper Series* 23617 (July).

Glaeser, Edward L. 2005. "The Political Economy of Hatred". *Quarterly Journal of Economics* 120, no. 1 (February): 45–86.

Golder, Matt. 2016. "Far Right Parties in Europe". *Annual Review of Political Science* 19:477–497.

Goodwin, Matthew, and Jocelyn Evans. 2012. "From Voting to Violence? Far Right Extremism in Britain". *Hope Not Hate:* 1–36.

Gould, Eric D., and Esteban F. Klor. 2016. "The Long-run Effect of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West." *Economic Journal* 126 (597): 2064–2114.

Graff, Peter. 2013. "Boston Suspect's Web Page Venerates Islam, Chechen Independence." *Reuters,* April 19.

Greenberg, Adam Eric. 2014. "On the Complementarity of Prosocial Norms: the Case of Restaurant Tipping During the Holidays". *Journal of Economic Behavior & Organization* 97:103–112.

Greenwald, Anthony G, and Mahzarin R Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes". *Psychological Review* 102 (1): 4–27.

Greenwald, Anthony G, Debbie E McGhee, and Jordan LK Schwartz. 1998. "Measuring individual differences in implicit cognition: the implicit association test." *Journal of Personality and Social Psychology* 74 (6): 1464.

Greenwald, Anthony G., and Thomas F. Pettigrew. 2014. "With Malice Toward None and Charity for Some: Ingroup Favoritism Enables Discrimination". *American Psychologist* 69 (7): 669–684.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-Term Persistence". *Journal of the European Economic Association* 14 (6): 1401–1436.

Haberman, Clyde. 1982. "Basic Fare Said to Barely Cover Taxis' Expenses." *New York Times,* March 23.

Haggag, Kareem, Brian McManus, and Giovanni Paci. 2017. "Learning by Driving: Productivity Improvements by New York City Taxi Drivers". *American Economic Journal: Applied Economics* 9, no. 1 (January): 70–95.

Haggag, Kareem, and Giovanni Paci. 2014. "Default Tips". *American Economic Journal: Applied Economics* 6 (3): 1–19.

Hall, Richard. 2013. "'Muslims Are Evil. Let's Kill Them All': US TV Commentator Erik Rush Provokes Furious Reaction With Boston Bombing Twitter Rants." *The Independent,* April 16.

Halla, Martin, Alexander F. Wagner, and Josef Zweimüller. 2016. "Immigration and Voting for the Far Right." Forthcoming, Journal of the European Economic Association, August.

Hanretty, Chris, and Nick Vivyan. 2015. "Estimating Constituency Opinion in Britain Technical Report."

Hensel, Alexander, Florian Finkbeiner, Philip Dudek, Julika Förster, Michael Freckmann, and Pauline Höhlich. 2017. "Die AfD vor der Bundestagswahl 2017: Vom Protest zur parlamentarischen Opposition". *Otto Brenner Stiftung Arbeitshefte* 91 (May).

Holmans, Alan. 2005. *Housing Statistics of Housing in Britain.* Cambridge Centre for Housing and Planning Research.

Inglehart, Ronald F., and Pippa Norris. 2016. "Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash". *Harvard Kennedy School Faculty Research Working Paper Series* RWP16-026 (August).

Jacquemet, Nicolas, and Constantine Yannelis. 2012. "Indiscriminate Discrimination: a Correspondence Test for Ethnic Homophily in the Chicago Labor Market". *Labour Economics* 19 (6): 824–832.

Jennings, M. Kent, Laura Stoker, and Jake Bowers. 2009. "Politics Across Generations: Family Transmission Reexamined". *Journal of Politics* 71 (3): 782–799.

Jha, Saumitra. 2013. "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia". *American Political Science Review* 107 (4): 806–832.

Kaltefleiter, Werner. 1966. *Wirtschaft und Politik in Deutschland. Konjunktur als Bestimmungsfaktor des Parteiensystems.* Köln: Westdeutscher Verlag.

Kennedy, Randy. 2001. "A Nation Challenged: Cabbies; Drivers Say They Risk Violence by Working, and May Even Lose Money." *New York Times,* September 24.

Kühnl, Reinhard, Rainer Rilling, and Christine Sager. 1969. *Die NPD. Struktur, Ideologie und Funktion einer neofaschistischen Partei.* Frankfurt am Main: Suhrkamp.

Lapiere, Richard T. 1928. "Race Prejudice: France and England." *Social Forces* 7 (1): 102–111.

Lee, Ulysses. 1966. *The Employment of Negro Troops.* Vol. 8. Government Printing Office.

Liepelt, Klaus. 1967. "Anhänger der neuen Rechtspartei: Ein Beitrag zur Diskussion über das Wählerreservoir der NPD". *Politische Vierteljahreszeitschrift* 8 (2): 237–271.

Little, K. 1998. *Negroes in Britain: A Study of Racial Relations in English Society*. Class, Race and Social structure. Routledge.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel. 2015. "Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT)." *American Economic Review* 105, no. 5 (May): 340–45.

Luttmer, Erzo F. P., and Monica Singhal. 2011. "Culture, Context, and the Taste for Redistribution." *American Economic Journal: Economic Policy* 3 (1): 157–79.

Lynn, Michael, and Michael McCall. 2000. "Gratitude and Gratuity: A Meta-analysis of Research on the Service-tipping Relationship". *Journal of Socio-Economics* 29 (2): 203–214.

Lynn, Michael, Michael Sturman, Christie Ganley, Elizabeth Adams, Mathew Douglas, and Jessica McNeil. 2008. "Consumer Racial Discrimination in Tipping: A Replication and Extension". *Journal of Applied Social Psychology* 38 (4): 1045–1060.

Madge, Charles, and Tom Harnett Harrisson. 1937. *Mass Observation.* Frederick Muller Ltd.

Malgouyres, Clèment. 2017. "Trade Shocks and Far-Right Voting: Evidence from French Presidential Elections". *Robert Schuman Centre for Advanced Studies Research Paper* 2017/21 (March).

Mathews, Kelly, Jessica Phelan, Nicholas Jones, Sarah Konya, Rachel Marks, Beverly M. Pratt, Julia Coombs, and Michael Bentley. 2017. *2015 National Content Test Race and Ethnicity Analysis Report*. United States Census Bureau.

McConnell, Brendon, and Imran Rasul. 2017. "Hispanic-White Sentencing Di¤erentials in the Federal Criminal Justice System." Unpublished.

Millgate, H.D. 2010. *Got Any Gum Chum?: GIs in Wartime Britain, 1942-1945*. History Press Series. History Press Limited.

Moore, B.L. 2013. "African Americans in the Military." In *The Columbia Guide to African American History Since 1939,* edited by R.L. Harris and R. Terborg-Penn. Columbia guides to American history and cultures. Columbia University Press.

Mudde, Cas. 2007. *Populist Radical Right Parties in Europe.* Cambridge, UK, and New York: Cambridge University Press.

Murray, Paul T. 1971. "Blacks and the Draft: A History of Institutional Racism". *Journal of Black Studies* 2 (1): 57–75.

New York Taxi and Livery Company. 2011. *Rule Book*.

———. 2014. *TLC Factbook*.

———. 2016. *TLC Factbook*.

Norris, Pippa. 2005. *Radical Right: Voters and Parties in the Electoral Market.* Cambridge, UK, and New York: Cambridge University Press.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa". *American Economic Review* 101 (7): 3221–3252.

Ochsner, Christian, and Felix Roesel. 2016. "Migrating Extremists". *CESinfo Working Paper* 5799 (March).

———. 2017. "Activated History – The Case of the Turkish Sieges of Vienna". Unpublished, *CESinfo Working Paper* 6586 (July).

Ockenfels, Axel, and Peter Werner. 2014. "Beliefs and Ingroup Favoritism." *Journal of Economic Behavior & Organization* 108:453–462.

Olson, L. 2010. *Citizens of London: The Americans Who Stood with Britain in Its Darkest, Finest Hour*. Random House Publishing Group.

Orwell, George. 1943. "As I Please". *Tribune* (December).

Pettigrew, Thomas F, and Linda R Tropp. 2006. "A Meta-Analytic Test of Intergroup Contact Theory". *Journal of Personality and Social Psychology* 90 (5): 751–783.

Pogue, F.C. 1954. *The Supreme Command*. United States Army in World War II: European theater of operations. Office of the Chief of Military History, Department of the Army.

Pope, Devin G, and Justin R Sydnor. 2011. "What's in a Picture? Evidence of Discrimination from Prosper. com". *Journal of Human Resources* 46 (1): 53–92.

Rash, F.L. 1942. *Internal Memo, Headquarters, Services of Supply, European Theatre of Operations*. National Archives reference: NAII; 290; 57/17/087; Box 43, File 218, 'Negro Troops', September.

Redvaldsen, David. 2016. "'Science must be the Basis'. Sir Oswald Mosley's Political Parties and their Policies on Health, Science and Scientific Racism 1931–1974." *Contemporary British History* 30 (3): 368–388.

Reynolds, D. 1995. *Rich Relations: the American Occupation of Britain, 1942-1945*. Random House.

———. 2006. *From World War to Cold War: Churchill, Roosevelt, and the International History of the 1940s*. OUP Oxford.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0 [dataset]*. Minneapolis: University of Minnesota.

Ruppenthal, R.G. 1978. *Logistical Support of the Armies: May 1941 – September 1944*. The European Theater of Operations: United States Army in World War 2 1. Department of the Army.

Sahner, Heinz. 1972. *Politische Tradition, Sozialstruktur und Parteiensystem in Schleswig-Holstein: Ein Beitrag zur Replikation von Rudolf Heberles "Landbevölkerung und Nationalsozialismus"*. Meisenheim: Hain.

Sasser, C.W. 2014. *Two Fronts, One War*. Frontline Books.

Saunders, Stephen G, and Michael Lynn. 2010. "Why Tip? An Empirical Test of Motivations for Tipping Car Guards". *Journal of Economic Psychology* 31 (1): 106–113.

Schaller, Bruce. 1993. "For-Hire Vehicle Fact Book".

———. 2006. "The New York City Taxicab Fact Book."

Schüller, Simone. 2016. "The Effects of 9/11 on Attitudes toward Immigration and the Moderating Role of Education." *Kyklos* 69 (4): 604–632.

Smith, Graham. 1987. *When Jim Crow Met John Bull: Black American Soldiers in World War II Britain*. I.B. Tauris.

Soltas, Evan J., and David Broockman. 2017. "A Natural Experiment on Taste-Based Racial and Ethnic Discrimination in Elections." Unpublished.

Spencer, I.R.G. 1997. *British Immigration Policy Since 1939: The Making of Multi-racial Britain*. Routledge.

Spenkuch, Jörg L., and Philipp Tillmann. 2017. "Elite Influence? Religion and the Electoral Success of the Nazis." Forthcoming, American Journal of Political Science.

Spolaore, Enrico, and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* 51 (2): 325–369.

Stanley, Damian A, Peter Sokol-Hessner, Mahzarin R Banaji, and Elizabeth A Phelps. 2011. "Implicit Race Attitudes Predict Trustworthiness Judgments and Economic Trust Decisions". *Proceedings of the National Academy of Sciences* 108 (19): 7710–7715.

Steinmayr, Andreas. 2017. "Exposure to Refugees and Voting for the Far-Right. (Unexpected) Results from Austria." Unpublished, LMU Munich.

Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence using Google Search Data". *Journal of Public Economics* 118:26–40.

Stouffer, Samuel A, Edward A Suchman, Leland C DeVinney, Shirley A Star, and Robin M Williams Jr. 1949. "The American Soldier: Adjustment During Army Life". 1.

Tabellini, Guido. 2010. "Culture and Institutions: Economic Development in the Regions of Europe". *Journal of the European Economic Association* 8, no. 4 (June): 677–716.

The Guardian. 2009. *Racism Inquiry May Bring Ban on BNP Membership for Teachers*. http://www.theguardian.com/education/2009/sep/30/ed-balls-orders-racism-inquiry.

The Spectator. 2009. *The Rise of British Racism May be Horribly Close*. http://www.spectator.co.uk/features/3648608/the-rise-of-british-racism-may-be-horribly-close/.

Uhlmann, Eric Luis, Anthony Greenwald, Andrew Poehlmann, and Mahzarin Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity". *Journal of Personality and Social Psychology* 97 (1): 17–41.

Vlachos, Stephanos. 2017. "Inglorious Wars and Political Radicalization." Unpublished.

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.

———. 2015. " Nazi Indoctrination and Anti-Semitic Beliefs in Germany". *Proceedings of the National Academy of Sciences* 112 (26): 7931–7936.

Waddell, S.R. 2010. *United States Army Logistics: From the American Revolution to 9/11*. PSI reports. Praeger Security International.

Widner, Daniel, and Stephen Chicoine. 2011. "It's All in the Name: Employment Discrimination Against Arab Americans." *Sociological Forum* 26 (4): 806–823.

Wilson, Valerie, and William M. III Rodgers. 2016. "Black-White Wage Gaps Expand With Rising Wage Inequality".

Winkler, Jürgen R. 1994. "Die Wählerschaft der rechtsextremen Parteien in der Bundesrepublik Deutschland 1949 bis 1993". In *Rechtsextremismus*, edited by Wolfgang Kowalsky and Wolfgang Schroeder, 69–88. Opladen: Westdeutscher Verlag GmbH.

Xu, Kaiyuan, Brian Nosek, and Anthony G Greenwald. 2014. "Psychology Data from The Race Implicit Association Test on the Project Implicit Demo Website". *Journal of Open Psychology Data* 2 (1).