Essays in Political Economy and Behavioral Economics

Anna Felicitas Becker

Submitted to the Department of Economics in fulfillment
of the requirements for the degree of

Doctor of Philosophy in Economics

University College London

June 2022

I, Anna Felicitas Becker, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Date: June 8, 2022

Abstract

This thesis uncovers and studies determinants of economic decision making and political attitudes that have so far received little attention in economics. It focuses on three factors - social image concerns, moral values and sexual competition - all of which are shown to be relevant for the functioning of communities. The first chapter studies whether public shaming played a role in men's decision to join the British Army during World War I by exploiting a natural experiment. At the beginning of the war, young girls would hand out white feathers to men not in uniform in an attempt to shame them as cowards. The chapter shows that this shaming strategy had an effect on recruitment numbers suggesting that social image concerns can induce costly altruistic behavior that benefits the group. Whereas the first chapter studies a factor inducing cooperation in societies, the second and the third chapter look at two different factors which can explain social and political disagreement. The second chapter explores how moral values shape beliefs about facts based on results from an online experiment. It shows how the salience of the moral dimension of a political debate increases polarization in beliefs between people on the left and the right of the political spectrum. The third chapter looks at the consequences a skewed sex ratio can have on the political preferences and xenophobic attitudes of young men. We hypothesize that in environments where male-male competition for female partners is high, the frustration from being single and concerns about status and male identity are more severe and can foster out-group hatred. Using observational data for Germany, this chapter provides evidence that in areas which have a significant surplus of men people are more likely to hold xenophobic attitudes and vote for right-wing extremist parties.

Impact Statement

This thesis studies how social image concerns, moral values and sexual competition shape economic decision making and political preferences. We provide empirical evidence which demonstrates how these factors affect individual behavior which is relevant for social cohesion and societal cooperation. These findings can be of interest for academics and policy makers alike. First, traditional approaches in economics typically consider these factors to be of only little relevance. Our results call that thinking into question. Second, our findings can shed some light on the conditions for cooperation within societies and between different cultures and can therefore be of interest to policymakers.

The first chapter studies the power of social image concerns. At the beginning of World War I, young women all over Britain would hand out white feathers to men not wearing a uniform yet to publicly shame them as cowards. I analyze whether this had an effect on recruitment numbers and find that the so-called White Feather Girls successfully shamed men into joining the army. The chapter therefore shows that social image concerns matter in high-stakes settings and can be a source for parochial altruism and large-scale cooperation. The study thus points to the functionality of social emotions such as shame and their role in sustaining cooperation. At the same time, it puts into question whether it can be justified to leverage them, for example in policy making, given that individuals are willing to pay such a high personal cost.

In the second chapter, we study whether moral values have an effect on individuals' beliefs about factual statements using a survey experiment. Orthodox Bayesian models in economics would typically assume that beliefs ("is" statements) can shape individuals' preferences about what ought to be, i.e. their values, but not vice versa. Our results, in contrast, document that values exert an influence on beliefs and that this effect is mediated by prior political leanings. This might have contributed to the polarization in the beliefs about facts observed in recent years. We can thus highlight a new and potentially very important channel for belief formation which has so far received little attention. In addition, this finding can provide insights to policymakers as to why individuals disagree more and more on objective facts. This can be particularly relevant, for example, for the communication of policy reforms to the

public.

The third chapter studies whether the lack of success in the partner market can explain xenophobic attitudes. We hypothesize that male-male competition can in parts explain the increase in anti-refugee hate crime and support for a new anti-immigrant right-wing party after a large influx of predominantly male refugees into Germany between 2015 and 2017. Our results suggest that anti-refugee sentiments are indeed more prevalent in areas where the share of women is low. This chapter can thus provide new insights into the social foundations of xenophobia. It also emphasizes a dimension of gender-specific migration that can be relevant for local policy making such as regional development and locally targeted labor market programs.

Contents

List of Figures

List of Tables

Acknowledgements

I am very grateful to the people who have advised and supported me while I was working on this dissertation.

I would like to thank my supervisors: Steffen Huck, for his encouragement and for his guidance when I would not see the wood for the trees; and Ian Preston, for his help and support in all situations. I would also like to thank Attila Lindner. I learned a lot from him as my advisor and his support and encouragement have been invaluable.

I thank Hans-Joachim Voth for his incredible mentorship. I would also like to thank him and Bruno Caprettini for their detailed feedback on the first chapter of this dissertation and for sharing their data.

I gratefully acknowledge financial support from the German National Academic Foundation, the German Academic Exchange Service (DAAD), the Economic and Social Research Council (ESRC) and the Social Science Research Center Berlin (WZB).

Lastly, a big thank you goes to my friends and colleagues on the PhD program and to all those who have supported me from near and far away throughout the whole time.

Introduction

Over recent years, we have been confronted with global and national challenges which are testing the resilience of our societies and political systems. Climate change, extreme inequality, public health crises, and refugee migration are just some of the issues causing a lot of tensions and political frictions. We also see that political opinions have polarized. People no longer only disagree about the best solution to a problem but also about the - often scientific - facts at the heart of it. This severely impedes consensus which is needed to tackle these problems. In addition, many of the challenges we are facing require individuals to pay a personal cost to contribute to the solution. However, it is not clear how to encourage such altruistic behavior when people cannot agree on a common goal.

This thesis seeks to better understand the reasons behind the polarization and discord we see in many societies today and factors that have led to large-scale cooperation and altruistic behavior of individuals in the past. It provides three studies which look into determinants of individual decision making and political preferences that have so far received rather little attention in economics. However, we see that many developments of the past years are difficult to explain with standard economics. This

thesis therefore aims to explore new approaches to uncover what it is that allows us to function as communities.

In Chapter 1 of this dissertation which is titled *"Shamed to Death: Social Image Concerns and War Participation"*, I study whether social image concerns have the strength to induce people to take risky altruistic actions for the benefit of their community. In this historical study, I look at Britain at the beginning of WWI. At that time, women all over the country would hand out white feathers to men not wearing a uniform yet. The white feather was well understood as a symbol of cowardice and was meant to shame these men into joining the British Army. Using an event study approach, I analyze whether the so-called White Feather Girls had an impact on recruitment and show that they were indeed successful with their strategy. Contemporary witness reports suggest that the receipt of a white feather was a painful experience for British men at the time and was perceived as an embarrassment. The study therefore provides evidence suggesting that social image concerns are an important and powerful driver of human behavior. It furthermore shows that individuals are willing to risk their lives in war in order to maintain a positive social image. Social image concerns can hence be functional as they induce behavior that is costly for the individual but benefits the wider community.

In Chapter 2, titled *"Motivated Political Reasoning: On The Emergence of Belief-Value Constellations"*, my co-authors Kai Barron, Steffen Huck and I analyze the causal relationship between moral values and factual beliefs. We think of moral values as "ought" statements, i.e. statements about how people would like the world to be when applying their own moral standards. Factual beliefs, on the other hand, which we also refer to as "is" statements, are statements about facts that can be

objectively verified. Orthodox Bayesian economics would predict that beliefs can exert an influence on values but not vice versa. We hypothesize, on the contrary, that beliefs about facts can also be influenced by personal values and that this is mediated by the political leanings of individuals. To test our hypothesis, we ran a large online experiment with more than 1,500 US participants. We strictly followed the analysis laid out in our preregistration. Our results show that beliefs are indeed influenced by moral values when we control for the political preferences of our participants. This indicates that values might have contributed to the observed polarization in beliefs about facts. We also show that subjects are not willing to distort their values in response to the financial incentives we provide. Moral values might therefore be a strong motivational force.

In the final Chapter 3, *"Social Foundations of Xenophobia"*, my co-author Hans-Joachim Voth and I study the relationship between skewed sex ratios and anti-refugee crime in Germany. For most people, finding a partner is an important objective in life. Several recent developments have, however, negatively affected the probability of finding a partner especially for young men. Gender-specific migration in East-Germany, which is often linked to the educational aspirations of women and poor labor market conditions, has left many regions with an immense surplus of men. We hypothesize that the resulting increase in male-male competition for female partners makes it more likely that men perceive foreigners as a threat. In our study, we look at Germany between 2015 and 2017 when a large influx of refugees was followed by an increase in anti-refugee hate crime and the support for an anti-immigration party. We provide evidence for a positive correlation between a surplus of young men and the occurrence of hate crime and anti-immigrant voting in different regions. We corrob-

orate our findings with an instrumental variable strategy which exploits the nearby availability of student places typically taken up by women. Our results indicate that the link between xenophobic attitudes and sexual competition needs further consideration as it has the potential to explain parts of the growing opposition to (refugee) migration in many countries.

This thesis exploits different settings and methods to better understand factors that can create unity and dissonance in communities. It suggests that there are factors which are often neglected in orthodox economic approaches that can have a meaningful impact on how we live together. Future work should hence test the external validity of these results and uncover further mechanisms to contribute to the solutions to some of the challenges ahead.

Chapter 1

Shamed to Death: Social Image Concerns and War Participation

Abstract

Can social image concerns cause people to take costly actions benefiting their community? Using newly collected data, I study the impact of public shaming on voluntary recruitment during World War I in England and Wales. At the time, young women in many towns and cities handed out white feathers to men in civilian clothes, marking them out as cowards. This was intended to encourage volunteering. I reconstruct a panel of "White Feather Girls" activity from local newspaper articles and exploit the staggered spread of the movement in an event study framework. Following episodes of public shaming, recruitment increased significantly: Volunteering surged by a third during the 10 days after the first mention of the White Feather Girls in the news. Confounding factors such as reporting of wartime events are unlikely to account for these patterns. These results suggest that public image concerns can have first-order effects on costly altruistic behavior that benefits the group.

1.1 Introduction

Most people care about what others think of them. In his *Theory of Moral Sentiments*, Adam Smith already argued that our ideas of both right and wrong as well as our actions reflect our nature as social beings. Recent theoretical work by Bénabou and Tirole (2006) emphasizes the importance of public perceptions, motivating people to engage in "honorific" actions. That social image concerns can shape decisions from education and voting to charitable giving has also been demonstrated empirically in several recent studies (Bursztyn, Egorov, and Jensen 2019; DellaVigna, List, and Malmendier 2012; Gerber, Green, and Larimer 2008). However, the stakes in most of these studies are relatively low. While it seems abundantly clear that people care about their image to some extent, it is an open question *how much* they care and whether such concerns can help to sustain large-scale cooperation in human societies (Gintis 2000; Grimalda, Pondorfer, and Tracer 2016).[1] Are social image concerns mostly minor, and only of empirical consequence in low-stakes settings, or are people willing to take major risks to preserve and enhance their public standing?

In this paper, I use newly-collected data from Britain during World War I to show that public shaming can lead to a particularly risky decision – volunteering for wartime service. Britain had an all-volunteer army until 1916, with 2.5 million men joining. As part of the recruitment effort, young women would approach men not wearing a uniform in the streets and present them with a white feather, a symbol of cowardice, in an attempt to make them join the Army (e.g. Gullace 2002). Using detailed newspaper reports on the activities of the "White Feather Brigade" and highly granular data on local military recruitment, I show in a difference-in-difference analysis that

1. Ager et al. (2021) argue that **personal rivalry** can induce risky actions. While also a form of public image concern, their setting is specific to closely knit groups of peers.

public shaming led to big increases in volunteering. These findings strongly suggest that social image concerns induced by the actions of complete strangers can lead to highly costly, altruistic behavior in a real-world setting.

The White Feather Girls, an informal movement of young women, emerged soon after the outbreak of World War I. The movement spread across the country and white feathers were handed out in all parts of the UK. There are no official records on the activity of the movement and no precise information on the number of members or incidents. Contemporary witnesses argue that the "idea [of the white feather] spread like a virulent disease" and that receiving a white feather caused "pain and acute embarrassment".[2] Because many men followed the call of the White Feather, several companies in war-relevant industries gave out badges to their employees, seeking to protect them from the shame of receiving a feather.[3] I compile the first comprehensive panel on the daily activities of the White Feather Girls, using newspaper records. I combine these data with newly-digitized historical records on daily recruitment in 121 cities in England and Wales during the first months of World War I, August 1914 to May 1915.

I examine the effect of the White Feather Girls on recruitment by exploiting the staggered reporting on the movement across cities, using an event-study design. In order to meet recent concerns regarding the standard implementation of event study designs using ordinary least squares, the approach is implemented using the imputation

2. IWM letter. Imperial War Museum staff, "Great War Index to Letters of Interest," n.d., Imperial War Museum, London.

3. See also Gullace (2002) for more details on the introduction of badges to protect workers. Figure 1.C.3, Panel (b), in the Appendix shows a newspaper article discussing the practice for workers from the R.A.F. (Royal Aircraft Factory) in December 1914. Figure 1.B.2 in the Appendix shows how the idea to introduce badges was discussed in the House of Commons in March 1915.

strategy proposed by Borusyak, Jaravel, and Spiess (2021). For each recruiting city, I look at the effect of the *first* mention as the point in time when the movement initially reached a city. I find a significant effect on volunteering in the ten days after a mention of the White Feather Brigade in the local newspapers. The effect is of considerable size. Volunteering is on average 36.6%[4] higher in the 10 days following the first mention of the White Feathers. The effect is not reversed in the subsequent weeks, i.e. it does not turn negative after the positive effect following treatment. This suggests that the White Feather Girls increased the total number of volunteers rather than making those who would have enlisted anyway bring their decision forward.

A key assumption in such a difference-in-differences style estimation is that in the absence of the reporting about the White Feather movement recruitment would have evolved in the same way in treated and untreated cities. I test the credibility of this assumption by showing that there are no significant deviations between the treated and the control cities prior to treatment – suggesting that the parallel trends assumption is satisfied. This suggests that the spread of the White Feather Movement was not a response to local volunteering trends. I can furthermore show that the results are robust to the inclusion of several important controls. First, I control for the daily number of casualties in each recruiting city. Second, I control for regional differences that might have influenced the decision to enlist in the Army, such as the share of self-employed and the share of those in paid employment as well as the share of females. The results remain largely unchanged. Including demographic controls, average recruiting is increased by 45%[5] in the ten days after activity of the White Feather Girls in comparison to average recruitment in the ten days before.

4. The standard error of this estimate is 0.135.
5. The standard error of this estimate is 0.1415.

I also control for region-specific time-varying news coverage as a potential confound. For instance, the mention of the White Feather movement in the local newspaper might coincide with differences in war reporting. To deal with this concern I apply textual analysis to a large corpus of 250,000 randomly selected articles covering the time period studied. I compute two measures for news coverage. First, I apply the Latent Dirichlet Allocation (LDA) to train a topic model on five topics. In addition, I compute a measure for negative sentiment using a lexicon that provides information on the emotional annotation of words. I compute the share of all articles in the vicinity of each recruiting city that refer to the different topics determined by the LDA as well as a measure for negative news sentiment. The inclusion of both of these controls does not alter the results. The overall effect on recruiting is still statistically significant and slightly higher than in the base specification.

Finally, I also show that the results are robust to alternative implementations of the event study. In particular, I show that excluding cities that were treated considerably later than the majority does not change the results. Lastly, I also re-estimate the results using the Generalized Synthetic Control Approach (Xu 2017) which allows me to relax the assumption of parallel trends but find results similar to the difference-in-differences approach.

The White Feather Girls anticipated that they would hit a nerve with young men in early 20th century Britain. But why was this shaming strategy so successful? As shown by Fehr and Gächter (2000) altruistic punishment can increase levels of cooperation in the public goods game. Individuals are willing to punish defectors even if this yields no personal benefits and is costly for themselves (Fehr and Gächter 2002; Fehr, Gächter, and Kirchsteiger 1997). Giving out white feathers to punish young

men not yet enlisted was indeed a costly action for these women. The public did not approve of the method deeming it not seemly and the girls also had to fear immediate retaliation by resentful men. These would often react violently to the receipt of a feather. Fehr and Gächter (2002) argue that negative emotions towards the defectors can explain why individuals are willing to pay a personal cost. Bowles and Gintis (2005) and also Gintis (2004), on the other hand, discuss the reactions of the defectors who receive the punishment. They present evidence suggesting that punishment functions as a signal of social disapproval which evokes emotions of shame in the defectors and which they seek to avoid by cooperating. By signaling their disapproval the White Feather Girls diverted the young men's focus away from the pride associated with joining the army to the shame of not serving their country.

War participation is strongly linked to traditional norms of masculinity. The narrative of the male war hero fighting for the country, receiving female adoration in return, is arguably as old as storytelling itself. Evolutionary psychologists believe that the greater variance in terms of reproductive success among men as compared to women has enhanced intra-sexual competition in men and can explain differences in physical aggressiveness between the sexes (Vugt, Cremer, and Dirk P Janssen 2007a). At the same time, men have incentives to form coalitions. In such a setting, a combination of intra-group cooperation and inter-group aggression, also known as "parochial altruism", may be an adaptive trait.[6] A potential explanation for the success of the White Feather Girls is therefore that they took advantage of men's primal instincts (Hart 2010).[7] In this sense, this study can be thought of as "existence result", offering

6. See e.g. Choi and Bowles (2007) and Bowles (2009).

7. In *The Mystery of Courage*, Miller (2002) argues: "With courage comes embedded a theory of manhood. In a significant number of cultures, as chastity was to women, so courage was to men: the virtue at the center of their gendered identity."

empirical support for anthropological theories arguing that males signal their fitness through war participation, and are perceived as sexually more attractive by females when they return as war heroes (Rusch, Leunissen, and van Vugt 2015).

This paper makes several contributions to the existing literature. First, I exploit a unique natural experiment to show that individuals care about their social image even when the stakes are extremely high. Previous research has shown that social image concerns can shape our behavior in many areas of life. These include for example education (Bursztyn and Jensen 2015; Bursztyn, Egorov, and Jensen 2019), career choices (Bursztyn, Fujiwara, and Pallais 2017), and credit card take-up (Bursztyn et al. 2017).[8] In these settings, however, the outcomes considered are typically relatively low cost. Here, I show that people would go as far as putting their lives on the line to maintain a positive social image and can thus demonstrate the power of social image concerns.

Second, this study provides important new insights on the functionality of social image concerns. The observation that people are willing to pay an extremely high personal cost raises questions about why humans have evolved to put so much weight on others' opinions – Adam Smith's question in the *Theory of Moral Sentiments*. Previous studies have shown that individuals are willing to take costly actions for the benefit of the group when they care about their relative standing among their peers. Ager et al. (2021), for example, study the behavior of German fighter pilots during WWII and find that pilots are more successful, but also run a greater risk of death, when a former peer received an honorable mention. Cantoni et al. (2019), Bursztyn et al. (2021) and Enikolopov, Makarin, and Petrova (2020) show that protest par-

8. For an overview see also Bursztyn and Jensen (2017).

ticipation in Hong Kong and Russia is affected by the behavior of friends and other peers. There currently is no observational evidence suggesting that social image concerns can foster pro-social behavior outside tightly-knit small groups. Results from both laboratory (e.g. Gächter and Fehr 1999; Grimalda, Pondorfer, and Tracer 2016) and field experiments (Gerber, Green, and Larimer 2008; Dellavigna et al. 2016; DellaVigna, List, and Malmendier 2012) suggest a relationship between social image concerns and altruistic behavior between strangers. The external validity of this existing evidence can be questioned because stakes are often low. This paper analyzes a unique real-world setting with high stakes to show that public image concerns can be important even in front of total strangers, leading to risky, altruistic actions.

The rest of this paper is structured as follows. In the next section, I will provide an overview over the historical background of WWI in the United Kingdom, the recruiting efforts and the White Feather movement. I will also introduce the data and the possibilities it provides for the analyzes as well as the limitations it sets. In Section 1.3, I introduce the empirical approach in detail and discuss the main findings as well as their robustness. Section 1.4 concludes.

1.2 Historical Background and Data

On the 4th of August, 1914 the UK declared war on Germany. It did so after Germany violated Belgian neutrality. At the outbreak of the war, the British Army was composed of 250,000 regulars. Half of them were serving overseas, and the Royal Navy absorbed much of the country's defense spending. The Army was much smaller than that of many other European countries. The French Army, for example, counted 1.3 million soldiers at the time, the German Army entered the war with a total strength

of 1.9 million. Britain also exclusively relied on professional soldiers at the time. Conscription was only introduced in 1916. During the initial months of World War I, the United Kingdom therefore stepped up its efforts to recruit volunteers.

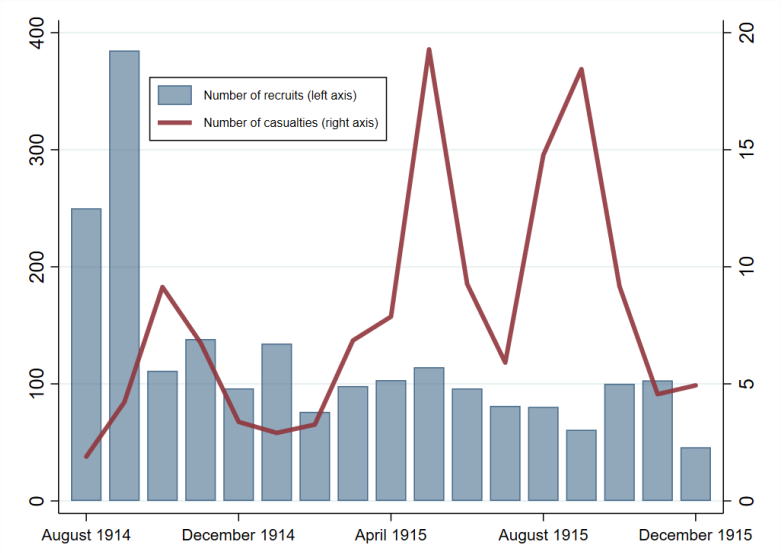
The Secretary of State for War, Herbert Kitchener, called for 200,000 volunteers in the first month of the war. The response was overwhelming: his call to arms was answered by around 250,000 volunteers in August and almost 400,000 in September. Figure 1.1 shows the number of volunteers per month from the beginning of the war until the introduction of conscription in January 1916. In the initial months, 2.5 million men enlisted voluntarily in the army; many of these joined during the first eight weeks of the war. As casualties mounted, the number of volunteers began to decline (see Figure 1.1, right axis). Nonetheless, "Kitchener's Army" consistently attracted around 100,000 volunteers each month between October 1914 and November 1915.

1.2.1 Why did British men enlist?

Why did British men volunteer in such large numbers? For many years, the idea that a wave of "war enthusiasm" swept the country following the declaration of war was common. It has more recently by questioned by historians (see e.g. Gregory 2008; Ferguson 1998). Whether the British public really thought that "the war would be over by Christmas" is equally doubtful. Hallifax (2010) identifies the huge amounts of money given to distress relief funds as one of several indicators suggesting that a rather long and destructive war was widely anticipated.

If it was not generalized nationalist feelings or false beliefs about the risks of joining the war, what can explain the large recruiting numbers? Ferguson (1998) suggests five motives. First, the efforts of the Parliamentary Recruiting Office (PCR), using all

Figure 1.1: Volunteering and Casualties per month (in thousands)



Note: The blue bars show the monthly number of recruits accepted for the Army and the Navy in the United Kingdom between August 1914 and December 1915 (left axis). The data has been digitized from NATS 1/399, National Archives (Kew). The red line shows the number of casualties per month obtained from the Naval & Military Press (right axis). All numbers are in thousands.

the tools of modern propaganda, might have had the intended effect on recruitment numbers. However, this cannot explain the surge in numbers right after the start of the war as the office only became operational from September onwards. Second, Ferguson highlights the importance of peer pressure. So-called "Pals" Battalions are considered important in getting friends and neighbors to join the army. The first of those, however, was also only founded in late August. Third, economic reasons and pressure from employers might have drawn men in. Some companies guaranteed jobs and pensions for volunteers. This may have boosted recruitment. However, Dewey (1984) shows that low wages are, if anything, negatively related to enlistment. Ferguson further highlights impulse as a potential driver. He cites Offer (1995) who suggests that a number of new recruits decided to join the war due to myopic

preferences, perceiving the train ride to the depot as a substitute for holidays.

Finally, Ferguson suggests that female pressure may have been important for the high recruiting numbers. Women were involved in the war effort in a variety of ways (Ward 2001), from fund-raising to volunteering as nurses. The decision whether or not to enlist increasingly became a matter of honor. Who would be better suited to put male self-esteem to the test than the opposite sex? The government later instrumentalized public image concerns in front of women to push men into the armed forces. Official recruiting posters proclaimed "Women of Britain say - GO!", or had the "women of London" ask: "Is your best boy wearing khaki? ... if your young man neglects his duty to King and Country, the time may come when he will neglect *you*."[9][10]

The original idea to use female pressure as a means to increase recruitment numbers, however, was introduced by Admiral Fitzgerald only weeks after the start of the war. It turned into a movement that spread quickly throughout the country.

1.2.2 The White Feather Girls

The White Feather movement was launched on August 30, 1914 when Admiral Charles Penrose Fitzgerald mounted the bandstand at Folkstone. Not satisfied with the recruitment efforts at the time and a convinced conscriptionist himself, he recruited about thirty women to hand out white feathers to mostly young men not wearing a uniform (for a detailed discussion of the phenomenon see e.g. Kilday and Nash (2017), Gullace (1997), and Gullace (2002)).

9. See Stevens (2016).
10. The latter poster can be found at: https://www.iwm.org.uk/collections/item/object/28305, accessed 31st of October, 2021.

In England, the white feather has been a symbol of cowardice since the late eighteenth century. The association is most likely related to cockfighting as cockerels with white feathers in their tails were considered a crossbreed inferior to pure-bred game cocks who do not exhibit such. In the 1902 novel "The Four Feathers" by A.E.W. Mason, a young military officer resigns from the army. Three of his comrades thereupon send white feathers to him. His fiancé also ends the engagement after confronting him with his cowardice, breaks a white feather from her fan, and hands him a fourth feather. Chastised, he returns to the army, fights in Sudan, and becomes a hero saving his unit from destruction. Thereupon, his fiancé takes him back.

The white feathers that Admiral Fitzgerald exhorted his followers to hand out were intended to do the same as Mason's fictitious feathers – to shame young men and humiliate them publicly, in front of females. Numerous women responded to his call and many more women all over the country imitated the original campaign in Folkstone. The movement, often referred to as the "White Feather Brigade" or "The Order of the White Feather", continued to exist long after conscription was introduced in 1916.

In March 1915, the practice was discussed for the first time in Parliament as concerns over the safety of non-military personnel were raised.[11] While no decision was reached at the time, the suggestion to distribute badges to individuals signaling that they are employed in industries of "national importance" and relevant for the war was later widely adopted to protect men not in uniform from being publicly shamed with a white feather. Men in civilian occupations would also receive letters and postcards with feathers, threatening them with making their "cowardice" public. The

11. See Figure 1.B.2 in Appendix for a transcription of the discussion in the House of Commons.

movement spread throughout the Empire (Stevens 2016).

There is no written evidence of how the movement was organized or how many members it counted. Nevertheless, its magnitude and effect must have been significant. A collection of letters archived at the Imperial War Museum in London bears witness to the campaign's impact. In 1964, the BBC issued a call in the national newspapers for "cowards" to step forward. They were looking for men who received a white feather as a symbol of cowardice in WWI as part of their research for a documentary about the Great War.More than 150 men replied with a letter to the BBC, sharing their memories fifty years after the war. The letters clearly show that a white feather did not leave the recipient cold. Instead, the men report how they reacted with great anguish and at times violence towards the White Feather Girls. Some men explicitly stated that receiving a white feather convinced them to go to the recruiting office the next day. Others provided detailed accounts of what prevented them from joining the army to explain why the white feather was unjustified in their case. There can be no doubt that the threat of receiving such a symbol of cowardice weighted heavily on the minds of some young men at the time:[12]

> Any young man who was not in Service uniform got a very cold shoulder from the women folk. Songs like "I don't want to lose you, but we think you ought to go" ... were rendered by women vocalists throughout the land... someone got the idea of the White Feather ... The idea spread like a virulent disease... But thank goodness, that tragic emblem of cowardice, the horrible white feather, never came my way.

12. From the letter from the BBC collection.

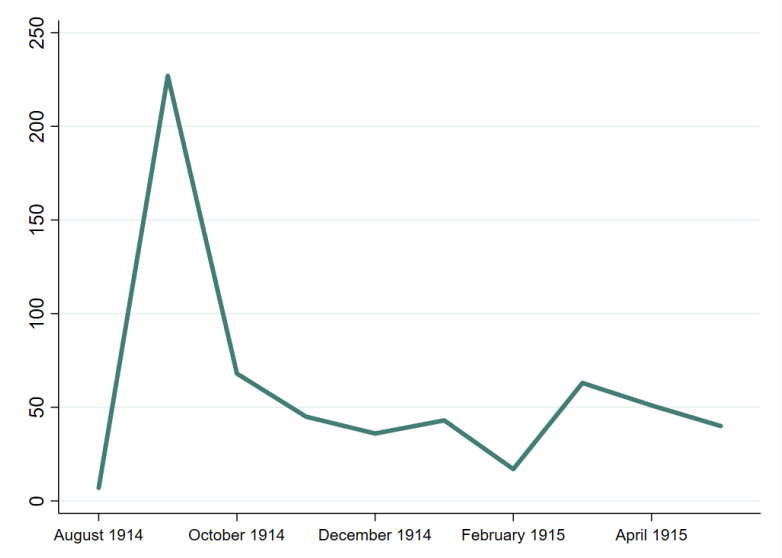
What is unclear is whether these recruiting efforts through public shaming had any effect. The same man quoted above said in his letter to the BBC that "it is doubtful if it [the white feather campaign] ever gained a recruit". Some historians have shared this skepticism. Robb (2014) considers the White Feather Girls "a minor patriotic outburst in the early months of the war", and argues that it failed because such public action was considered "too unladylike". The movement was also critized at the time for harassing men and increasing cleavages in society. The famous playwright George Bernard Shaw publicly critized the white feather girls as early as September 1914. In order to accurately assess the impact of the white feather girls, we require quantitative data on their activities at high frequency.

1.2.3 Quantifying the White Feather Movement: News from Home and from the Front

Since the White Feather Girls were not a centrally organized movement, there are no official records of members and campaigns. While there are recollections of individuals and diary entries providing details of specific events, we have no complete list of the dates, places or number of incidents. However, news of the White Feather Girls and their activities spread quickly throughout the United Kingdom. Local and regional newspapers reported their activities. They tracked incidents not only at the beginning, when the first feathers were handed out, but also during later months of the war as Figure 1.2 shows. The movement was discussed in letters to the editor and in various comment sections. Cartoons were published ridiculing the women involved in the movement and entire pages were dedicated to discussions of whether a decent woman would engage in such behavior. Other reports on the movement were more positive (see e.g. the newspaper article shown in Figure 1.C.3 in the Appendix).

In total, I found 597 articles which cover the White Feather Movement between August 1914 and May 1915, the time period this study covers. About one third of the articles were published in September 1914, right after the movement was initiated in Folkstone. Afterwards, around 50 articles were published each month.

Figure 1.2: Articles on the White Feather Girls



Note: The graph shows the monthly aggregated number of articles that refer to the White Feather movement found in the British Newspaper Archive for the period between August 1914 and May 1915.

During the war, newspapers were the main source of information. The first radio station started broadcasting only after the war. The local and regional papers played a special role as they provided a form of contact with the loved ones at the front. In the early years of the war, extracts from letters home from soldiers at the front were published in the local papers. In addition, the local papers provided detailed accounts of the number of casualties and were therefore highly anticipated by the local population and read with great care.

1.2.4 Data on Recruitment and White Feathers

To analyze the effect the White Feather Girls had on recruiting in England and Wales at the beginning of WWI, this study draws mainly on two data sources. First, I use data on daily recruitment in 121 recruiting cities in England and Wales. The data was digitized from original recruitment records kept at the National Archives (Kew).[13] The daily data covers voluntary recruitment between August 1914 and May 1915.

As there is no official account on the number of White Feather incidents, I use information from local newspapers. For that purpose, I extracted all relevant articles in the British Newspaper Archive (BNA). The British Newspaper Archive works in partnership with the British Library to digitize their large collection of local and regional newspapers. All articles found relating to the White Feathers were carefully read and checked for relevance. The articles were then located using the publishing place of the newspaper. In total, the BNA has digitized articles from 253 local newspapers in England and Wales for the time. Out of these 253 newspapers, 181 have published one or more articles on the White Feather movement.

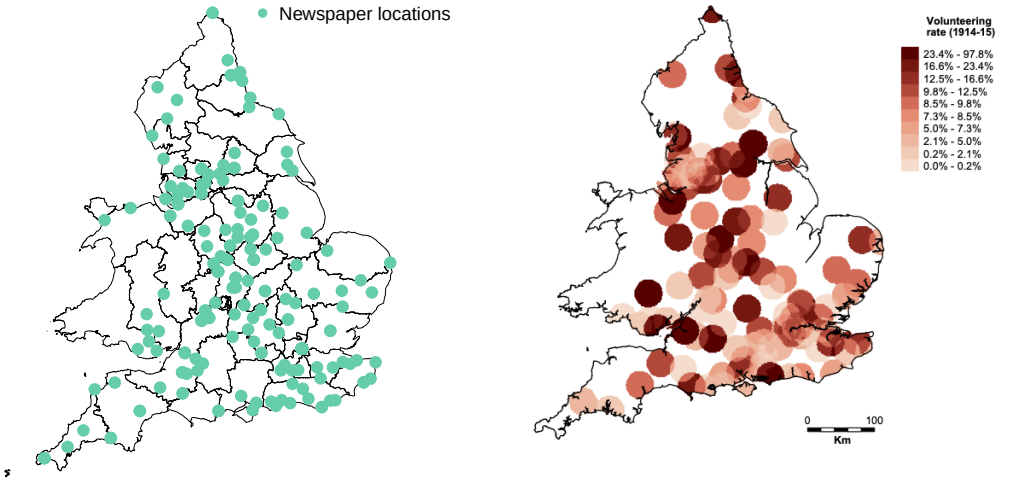
The left panel of Figure 1.3 shows the locations of newspapers that reported on the activities of White Feather Girls.[14] The right panel shows the aggregate recruiting numbers in the 121 recruiting cities in England and Wales for which daily data is available. The number of volunteers is normalized by the number of available recruits in the different cities. This information is obtained from the national census of 1911.

13. While the data do not cover all volunteers who signed up in England and Wales at the time it follows the same trend as overall recruiting numbers (see Figure 1.A.1 in the Appendix).

14. Figure 1.C.4 in the Appendix shows in addition which newspapers publish articles on the White Feather Girls and where most articles on White Feather Girls are published.

The information is available at the level of parishes. To create a meaningful reference for the different recruiting cities, each parish in the data was aggregated according to its closest recruiting city.

Figure 1.3: White Feathers and Volunteering



(a) Publishing Cities **(b)** Volunteering in Recruiting Cities

Note: Panel (a) shows the places in England and Wales in which a local newspaper was published. The data comprises all local newspapers found for the period between August 1914 and May 1915 in the British Newspaper Archive and use the places of publication provided there. Panel (b) on the right shows volunteering rates in the 121 recruiting cities for which daily volunteering rates are available aggregated for the period between August 1914 and May 1915. The data has been digitized from historical tables kept at the National Archives (Kew) (NATS 1/398).

To control for general sentiment in the news I also use a representative sample of articles from the British Newspaper Archives. The articles were randomly chosen from the available newspapers taking the periodicity of newspapers into account and ensuring that a similar number of articles was sampled for each week covered by the available recruitment data. The text of the scanned newspaper articles has been digitized using optical character recognition (OCR) by the BNA and is therefore available

in a machine-readable format. This allows me to create a data set combining the place and date of publication with the text of each article.

1.3 Empirical Strategy and Results

1.3.1 Treatment Assignment

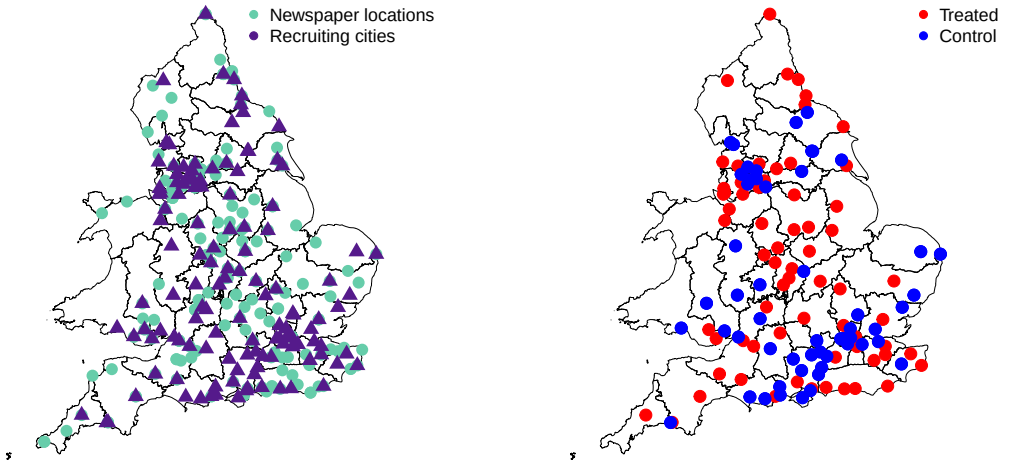
To show the effect of the White Feather Movement on recruitment in England and Wales I make use of a difference-in-differences design with staggered adoption of treatment, exploiting both the regional variation in the publication of White Feather articles and the variation over time.

Using daily data, I look at a high-frequency outcome for recruitment. I assign treatment in a way that allows me to study the most immediate effect of articles covering the White Feather Movement on volunteering to fight. This assignment represents the most conservative way of assigning articles to recruiting cities, in the absence of circulation information for local newspapers.

Figure 1.4 shows the locations of newspapers and the locations of recruiting cities in the left panel. Since there is no information on either the circulation or on the area of distribution for the newspapers, I assign each article published on the White Feather to the *closest* recruiting city. The idea is that a man reading about the White Feather Girls would go to the nearest recruiting city in the data set to join the Army. A city is then considered treated when the first White Feather article has been published. Untreated or control cities are hence those cities that have never been closest to any newspaper location publishing an article on the White Feathers.

The right panel of Figure 1.4 shows the resulting pattern of treated and control cities

Figure 1.4: Assignment of Treatment



(a) Recruiting and Publishing Cities **(b)** Treated and Control Cities

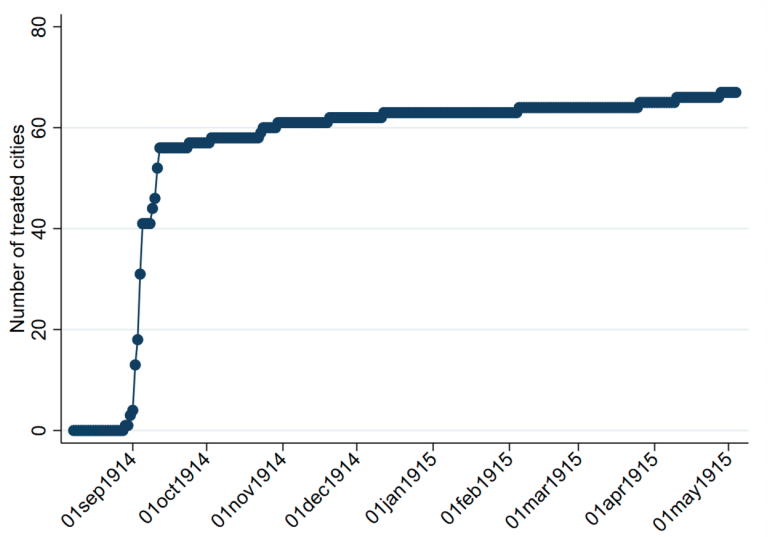
Note: The blue dots in Panel (a) mark the places in England and Wales in which a local newspaper was published. The data comprises all local newspapers found for the period between August 1914 and May 1915 in the British Newspaper Archive and use the places of publication provided there. The purple triangles in Panel (a) show the locations of the 121 recruiting cities for which we have daily recruitment data. Panel (b) shows how recruiting cities were divided into treated and control cities according to the treatment definition. A city is considered treatment after it has been the closest city to a newspaper publishing an article on the White Feathers.

for the sample. In total, there are 67 cities that receive treatment whereas the remaining 54 cities remain in the control group (never-treated). The majority of cities received treatment in the first half of September 1914, with only 10 cities subsequently being treated. Table 1.D.1 in the Appendix shows that the treated and control cities are strongly balanced in terms of observable characteristics. The difference in the share of females is significant on the 10% level but smaller than 0.2% of the mean in the control group. The difference in the share of people employed as “crafts and related trades workers” is significant but small in magnitude, too.

Apart from the regional variation in exposure to news about the White Feather Girls,

the empirical strategy also exploits the staggered publication of these articles. As Figure 1.5 below shows, the majority of cities were treated right when the movement began, i.e. in the first half of September 1914. Only four cities received treatment according to the definition applied here in 1915.

Figure 1.5: Change in Treatment Status over Time



Note: The graph shows the number of treated cities over time, i.e. for each day covered by the data it shows the number of cities who have been closest to the place where an article on the White Feather Girls has been published. Only the first article published in each location on the White Feather Girls is taken into account.

1.3.2 Estimation Approach

To identify the causal effect of the White Feather movement on volunteering at the beginning of WWI, I rely on a difference-in-differences approach. This approach looks at the effect on the first days after initial treatment separately. More specifically, I assume that the outcome of interest, the volunteering share V_{it}, can be described

by the following equation:

$$V_{it} = \alpha_i + \alpha_t + \tau_{it} D_{it} + \varepsilon_{i,t}. \tag{1.1}$$

The volunteering share V_{it} is defined as the daily number of recruits in city i at day t divided by the number of eligible men in and around a recruiting city. The number of recruits has been digitized from archival data as described in Section 1.2.4 and the number of eligible men has been calculated from 1911 census data. Eligible men comprise all non-disabled native-born men who are at least 13 and at most 50 years old. Census data is available at the level of parishes. To obtain relevant numbers for the recruiting cities in the data, every parish has been assigned to the nearest recruiting city. The volunteering share V_{it} is defined in percent.

In the equation above, α_i describes city fixed effects, α_t captures time fixed effects. D_{it} indicates treatment with $D_{it} = 1[t \geq E_i]$ where E_i is the day when a city receives treatment and $\varepsilon_{i,t}$ is the error term with $E[\varepsilon_{i,t} | \alpha_i, \alpha_t, D_{it}] = 0$. We are interested in the treatment effect τ_{it} which captures the impact of initial reporting on the White Feathers on volunteering. The approach assumes parallel trends, i.e. that treated and untreated cities would have followed the same volunteering trend in the absence of treatment.

In a conventional approach, the above would be estimated using an event-study design in which the outcome is regressed on time and unit fixed effects as well as both leads and lags of the treatment using ordinary least squares. As recently shown by Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021) and Strezhnev (2018), however, the estimands of such an approach are not reliable and do not pro-

vide the causal effects of interest even under random assignment of treatment.[15] The bias is driven by comparisons of earlier-treated with later-treated units when there are heterogeneous treatment effects.[16] In the setting considered here, a bias could for example arise if individuals treated later have a better understanding regarding the risks of participating in the war, e.g. by observing the outcome of large battles. This study therefore uses an imputation estimator which takes the bias into account.

Several robust estimators have been proposed to deal with this bias (see for example Chaisemartin and D'Haultfœuille (2020), Sun and Abraham (2021) and Callaway and Sant'Anna (2020))[17]. This study makes use of the imputation estimator as proposed by Borusyak, Jaravel, and Spiess (2021). Their estimator is constructed in three steps. First they estimate the unit and day fixed effects from equation 1.1 using only untreated observations. The estimates are then used in the second step to obtain unbiased estimates of the treatment effects $\hat{\tau}_{it} = V_{it} - \hat{\alpha}_i - \hat{\alpha}_t$. While these day-unit treatment effects cannot be estimated consistently, they show that consistent estimates can be obtained for averages of many observations. In the final step the average

$$\hat{\tau}_h = \frac{1}{|I_h|} \sum_{i \in I_h} \hat{\tau}_{i,E_i+h} \tag{1.2}$$

is therefore computed for the h days since treatment where I_h is the set of cities observed in period $E_i + h$.

Their approach is particularly appealing in this setting since the application is straightforward and allows for a flexible inclusion of both time-variant and time-

15. See also Sun and Abraham (2021), Callaway and Sant'Anna (2020), Imai and Kim (2021), Borusyak and Jaravel (2017) and Athey and Imbens (2018).
16. See Baker, Larcker, and Wang (2021), for a survey of the literature.
17. See also Cengiz et al. (2019) for an example of a stacked regression approach.

invariant control variables which speaks to the available data. It also allows easy assessment of whether the identifying assumption of parallel trends holds. Unlike in a conventional event study approach, pre-trends are not estimated together with the treatment effects but separately by estimating the following regression for untreated observations only:

$$V_{it} = \alpha_i + \alpha_t + \sum_{p=-P}^{-1} \gamma_p 1[t = E_i + p] + \varepsilon_{it} \tag{1.3}$$

where $1[t = E_i + p]$ are variables indicating if a unit is treated 1 to P days later. The pre-trend coefficients can be computed simultaneously with the treatment effects although the effects are estimated separately rather than jointly which has further attractive properties.[18]

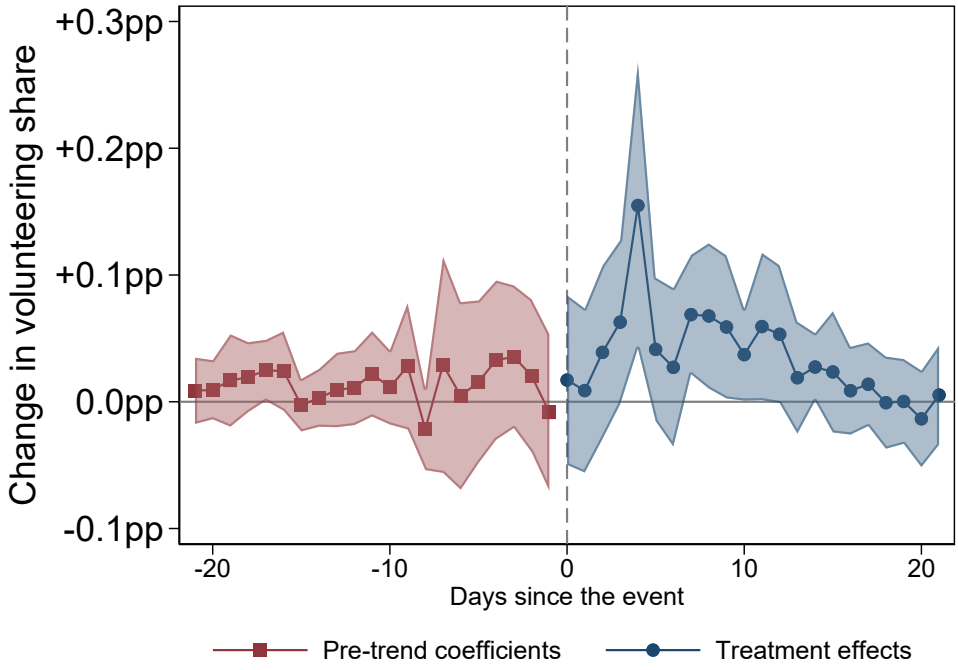
1.3.3 Main Results

The articles collected on the White Feathers record either concrete incidents or report on the movement more generally. In cases where events are reported it must have taken at least one day before the report of the event can be found in the local news. To take this into account, the treatment indicator is redefined to switch from zero to one two days before treatment.

Figure 1.6 shows the results from the estimation of the treatment effects as defined in equations 1.1 and 1.2 (blue dots) and the results from the estimation of the pre-trends as defined by equation 1.3 (red squares). The shaded areas indicate the 95% confidence level around the estimated coefficients.

18. For a detailed discussion of the approach and an application to school closures on the transmission of COVID-19 see Von Bismarck-Osten, Borusyak, and Schönberg (2021).

Figure 1.6: Effect of White Feather Activity on Volunteering



Note: The graph shows the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample comprises all 121 recruiting cities. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of eligible men from the 1911 Census.

The pre-trend coefficients are all close to zero and statistically insignificant. They do not exhibit a clear upward or downward trend. The F-test accepts the hypothesis that the pre-trend coefficients are jointly equal to zero with a p-value of 0.07. This supports the identifying assumption of parallel trends.

The treatment effects are significantly different from zero on the 5% level and positive on days three, four, seven to eleven and fourteen after treatment. This therefore suggest that the White Feathers increased the volunteering rates during the early days

of the war. The average volunteering rate in the 10 days before treatment is 0.163%. The effect is largest on day four or – keeping in mind that the treatment indicator switches to from zero to one two days before the date of publication of an article – two days after article publication. On that day, the increase in the volunteering share due to reporting on the White Feathers Girls is 0.155 percentage points. This implies that two days after a local paper reports on the movement, the volunteering share almost doubles in the treated cities.

The average treatment effect over the first ten days after publication of an article is 0.06 percentage points. This implies that in comparison to the ten days before treatment the average daily volunteering share increases in the treated cities by 36.6% for about ten days after an article is published.[19] This corresponds to approximately 33,000 soldiers which is equivalent to 6.26% of the recruits who have in total signed up in August, September and October 1914.[20]

Note that the effect is not reversed on subsequent days and in subsequent weeks. Figure 1.E.5 in the Appendix shows the treatment effect in the two months after treatment and also indicates no reversal over this longer period. This suggests that the men who volunteered because of the White Feather Girls would not have done so otherwise, i.e. this suggests the White Feather Girls did not just bring the decision to volunteer of these men forward.

19. The standard error for this estimate is 0.1350.
20. These numbers are based on a back-of-the-envelope calculation.

1.3.4 Robustness

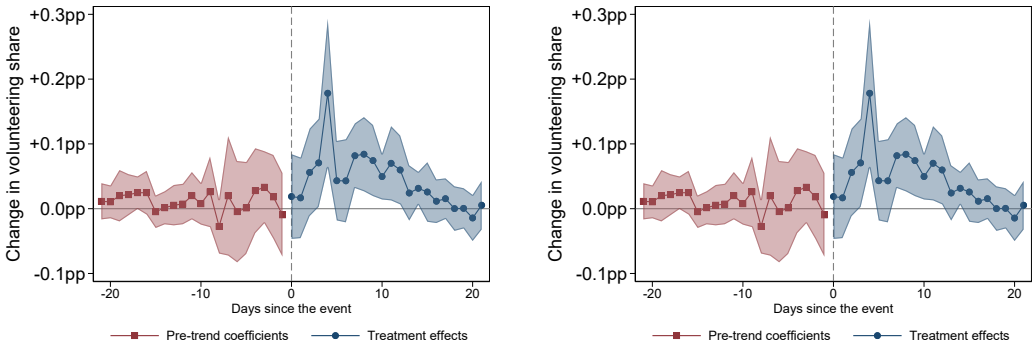
1.3.4.1 Including Controls

In this section, I make use of a number of control variables that I have not taken into account in the previous analysis. The previous estimation takes into account constant characteristics of locations that could be relevant for recruitment. A potential concern could be that fixed characteristics of the cities have a time-varying component that affects volunteering. To meet this concern I use time-invariant information from the 1911 census and include the variables as interactions with period dummies. I include the share of females to deal with any potential concerns arising from the small but statistically significant difference between the treated and the control group as seen in Table 1.D.1. Furthermore, I include the share of the population born in another country as national identity might be an important determinant in the decision to go to war as well as a measure for the share of young individuals in the population. To control for economic factors that might impact volunteering rates, I include the share of self-employed, the share of farmers, and the share of those in paid employment.

As Figure 1.1 suggests recruitment might be (negatively) related to casualties. I therefore control for the number of dead soldiers. Since I have information on the date of death and the residence of those soldiers I can assign them to the recruiting cities in the data. The results using census controls and a control variable on the number of casualties are reported in Figure 1.7. The left panel shows results including the above mentioned information from the 1911 census, the right panel controls for casualties in addition to that.

As the figure shows, the results are robust to the inclusion of those controls. Again,

Figure 1.7: Effect of White Feather Activity on Volunteering (including Controls)



(a) Including Census Controls

(b) Including Census Controls and Control on Casualties

Note: The graph shows the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample comprises all 121 recruiting cities. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of recruitable men from the 1911 Census. Controls from the Census comprise the share of females, the share of the population born in another country, the share of young individuals in the population, the share of self-employed, the share of farmers and the share of those in paid employment. Casualties are the number of soldiers who died each day whose residence is in an area assigned to the respective recruiting cities. Information on soldiers who died come from the Naval & Military Press.

none of the pre-trend coefficients is significantly different from zero and they do also not indicate a clear trend suggesting that the parallel trend assumption continues to hold. The general pattern for the daily treatment effects is also largely similar to the main results discussed in the previous section. The effect of White Feather Girls activity on volunteering is significantly different from zero on days three and four as well as on days seven to twelve and fourteen after treatment when census controls are included as in Panel (a) of Figure 1.7. The effect is slightly larger and estimated with more precision. On average, the volunteering share is 0.06 percentage points higher in the 11 days after there has been an article on the White Feather Girls. In

comparison to the average in the ten days before treatment, recruitment was therefore 45%[21] higher. Adding the number of soldiers who died (in logs) as a further control does not change the results notably as shown in Panel (b) of Figure 1.7.

1.3.4.2 Controlling for News Sentiment

A further concern could be that the articles on the White Feather movement are just picking up other trends of war reporting in the newspapers. Using the random sample of articles from the BNA corpus, I therefore compute two measures to control both for the topics covered by the news as well as for the sentiment of the news.

I have obtained approximately 250,000 articles from the British Newspaper Archive covering the period at the beginning of WWI for which I have information on recruitment. They were randomly selected from the overall BNA corpus making sure that the different periodicity of the newspapers is reflected. In a first step, I create a topic model. The idea of topic modeling is to classify documents or, as is the case here, articles into natural groups in an unsupervised way. One of the most common algorithms used for topic modeling is the Latent Dirichlet Allocation (LDA). It assumes that every document is a mixture of topics and that every topic is a mixture of words. The LDA estimates both the mixture of words and the mixture of topics at the same time. In a first step, I have trained the algorithm on a 50% random sample of articles of the 250,000 articles in the data set on a five-topic model.

The five topics with the 25 most important terms by topic can be found in Figure 1.8 below. The y-scale depicts the per-topic-per-word probabilities, β, from the model. As was expected, the war plays an important role in newspaper reporting during

21. The standard error of this estimate is 0.1415. Average recruiting in the eleven days before is 0.159%.

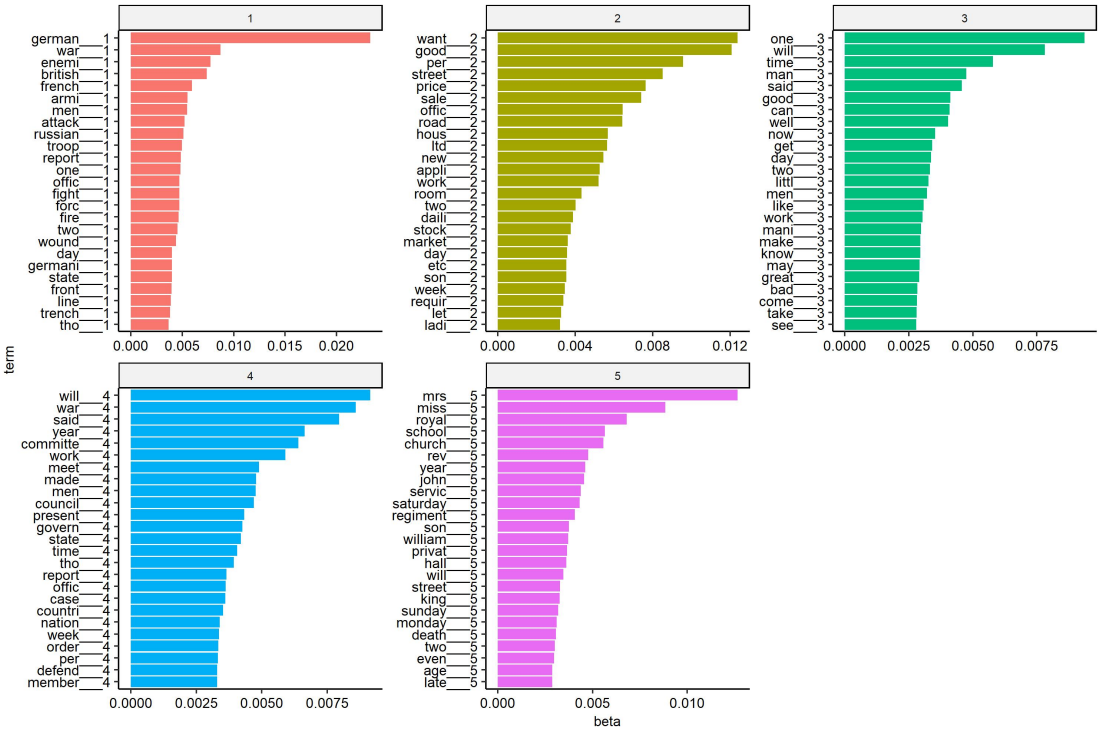
the time. Topic 1 appears to be a representation of reporting from the front and recent war events with the most important words being “german”, “enemi” and “war” itself.[22] The war is also important in topic 4 but unlike in topic 1 it is rather a subject of politics as the topic features terms such as “committee”, “council”, “state” and “govern” otherwise. Topics 2 and 3 are unrelated to the war. The second topic does not show a relation to war topics but covers the economy featuring words such as “good”, “price”, “sale” and “market”. The third topic cannot easily be characterized.

Topic 5 indirectly relates to the war covering the reporting of casualties with words like “church”, “death” and several common names such as “john” and “william” among the most important words. Including a control for topic 5 can thereby also cover potential concerns that it is not the date of the casualties which matters for enlistment but the day when the casualties are reported in the news. For strategic reasons and potential delays in news transmission during the war, more or less time might have passed between those dates. I can therefore provide an alternative approach to account for the effect of casualties on volunteering in the estimation.

In a second step, I have then used the remaining half of articles in the data set and applied the algorithm to classify the articles into one of the five topics. For each article, I have computed the probabilities to fall into each of the five topics and assigned it to the topic it is most likely to belong to. For every recruiting city and every day the data covers I have then looked at the articles published by newspapers within a radius of 50km. Finally, I have computed the share of articles falling into each of the five topics.

22. Word stems were used in the analysis and are also shown here.

Figure 1.8: Latent Dirichlet Allocation (LDA) on Articles from Local Newspapers



Note: The Latent Dirichlet Allocation (LDA) was trained on a random sample of 125,000 articles from our sample of local newspaper articles collected from the British Newspaper Archive. The Figure depicts the five topics with the 25 most important terms by topic. The y-scale depicts the per-topic-per-word probabilities, β, from the model.

Apart from the topics covered by the news, it is also possible that the news sentiment contributed to the decision to volunteer. The tone with which the news is spread could potentially matter just as much as the content or at least shape how the content is perceived. To characterize the news in terms of their sentiment, every word in every article was categorized using the NRC Word-Emotion Association Lexicon. The lexicon consists of a list of English words and their association with eight basic emotions which are anger, fear, disgust, surprise, anticipation, sadness, joy and trust. The annotations were done manually using crowdsourcing (for details see Mohammad and Turney (2013)).

Rather than using all eight categories, I have categorized words according to whether they were annotated with a negative sentiment (anger, fear, disgust, sadness) or with a positive sentiment (surprise, anticipation, joy and trust). This has been done to avoid considerable miss-classification as many words are associated with more than one emotion. For each recruiting city, I have pooled the articles published with a radius of 50 km each day and computed the number of words with a negative sentiment as a share of all words. This serves as an indicator for negative news sentiment in the vicinity of each recruiting city.

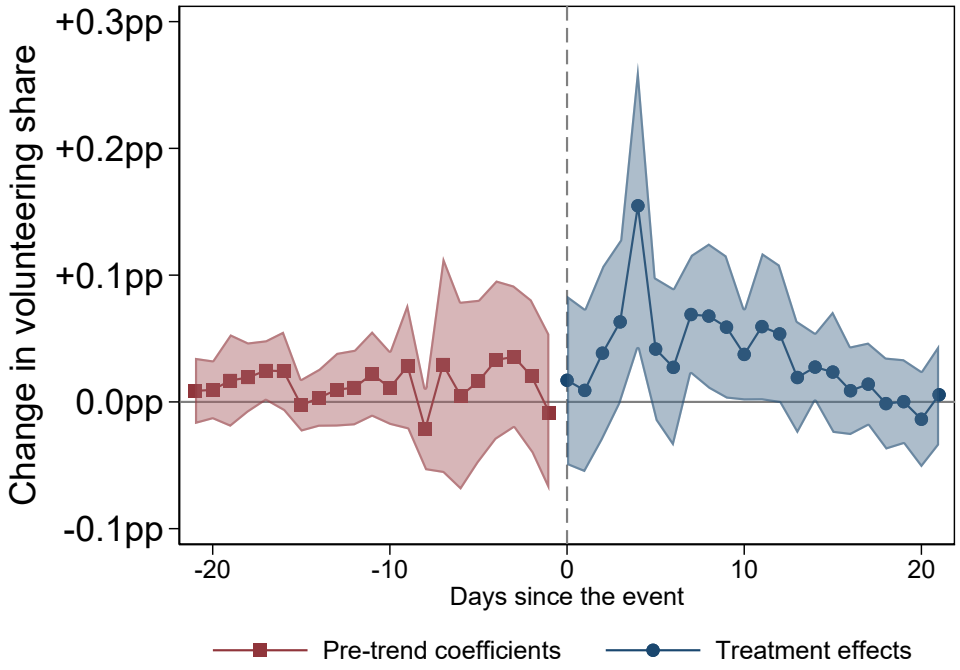
Figure 1.9 shows the imputation results including the controls both for the topics covered as well as for the sentiment of the news. In general, the results are very similar to the specification without controls depicted in Figure 1.6. The pre-trend coefficients are all close to zero and statistically insignificant. As before, we see a positive treatment effect in the days after treatment which is similar in magnitude to the previously estimated effect. The average effect over the ten days after a White Feather article has been published is 0.06 percentage points which implies that volunteering increased on average by 36.6%.[23][24][25]

1.3.4.3 Excluding the Late Treated

While the majority of recruiting cities have been treated soon after the White Feather movement was initiated, some cities are treated only later. To test whether the results are in some way driven by these late treated units, I restrict the sample in two steps.

23. The standard error for this estimate is 0.1356.
24. The effect is 0.1 percentage points larger than in the baseline specification which is not using any additional controls.
25. Figure 1.E.6 in the Appendix shows the results including census controls, controls for the number of soldiers who died and controls for news sentiment. Again, news sentiment does not change results much in comparison to those presented in Figure 1.7.

Figure 1.9: Effect of White Feather Activity on Volunteering – Controlling for News Sentiment

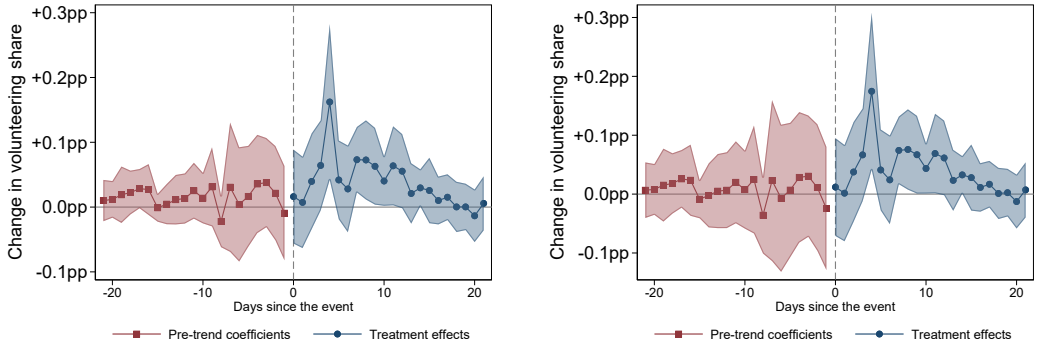


Note: The graph shows the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample comprises all 121 recruiting cities. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of recruitable men from the 1911 Census. Controls include the daily share of articles within a radius of 50km around the recruiting cities falling into each of the five topics defined by the LDA as depicted in Figure (1.8). Topic 3 serves as the reference. I also control for negative sentiment measured as the daily share of articles dominantly associated with a negative emotion (anger, fear, disgust and sadness) within a 50km radius around the recruiting cities.

First, I exclude all cities that were treated in 1915 only and in a second step I also exclude cities that have been treated after September 1914. The results are depicted in Figure 1.10 in the left panel for the sample comprising only cities treated in 1914 and in the right panel for cities treated in August and September 1914. In both

cases, the treatment effects are not different from those estimated based on the full sample. We see that the pre-trend coefficients are estimated with less precision but still insignificant on the 95%-level.

Figure 1.10: Effect of White Feather Activity on Volunteering – Restricting the Sample



(a) Exclude late treated (1915)

(b) Exclude late treated (after September 1914)

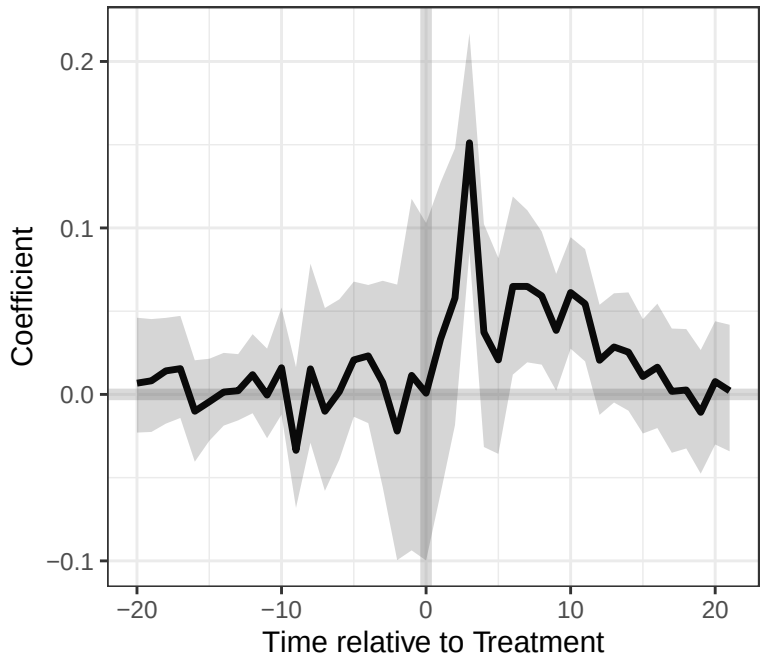
Note: The graphs show the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample for Panel (a) comprises 117 recruiting cities as it excludes those cities that were treated only in 1915. Panel (b) comprises 111 cities as it also excludes those treated after September 1914. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of recruitable men from the 1911 Census.

1.3.4.4 Generalized Synthetic Control Approach

The previous results strongly indicate that the assumption of parallel trends in the difference-in-differences approach is satisfied and that the estimated treatment effects therefore are identified. Nevertheless, I present results based on the Generalized Synthetic Control (GSC) method which allows me to relax this assumption. The Generalized Synthetic Control method proposed by Xu (2017) and Liu, Wang, and Xu (2020) uses the control group data in a first step to estimate an interactive fixed effect model where unit-specific intercepts are interacted with time-varying coeffi-

cients. The latter, also referred to as latent factors, are then used to compute factor loadings for each treated unit by using only pre-treatment outcomes. In the final step, the treated counterfactuals are imputed based both on the estimated factors and the factor loadings. The GSC approach can thus generalize the synthetic control approach as it allows for multiple treated units and variable treatment periods.[26]

Figure 1.11: Effect of White Feather Activity on Volunteering – Generalized Synthetic Control Approach



Note: The figure provides estimation results based on the Generalized Synthetic Control (GSC) Method proposed by Xu (2017). Panel (a) shows the treated (black line) and the control averages (blue line) for three weeks (21 days) before and after the day of treatment. Panel (b) shows the estimated treatment effect and the pre-trend estimated, again for three weeks (21 days) before and after treatment occurs. They gray-shaded areas indicate confidence intervals based on bootstrapped standard errors. The sample comprises all 121 recruiting cities. In both panels, the y-axis indicates the volunteering rate in percent.

Figure 1.11 shows the estimated treatment effects using GSC which support our previ-

26. For specific weights, the GSC approach coincides with the approach of Borusyak, Jaravel, and Spiess (2021).

ous findings. There are no significant differences in the pre-trends prior to treatment with an exception in day 9 before treatment occurs. The effect is then statistically significant and positive in the first 11 days after treatment. The overall effect is of very similar magnitude as the results from the imputation method proposed by Borusyak, Jaravel, and Spiess (2021). The generalized synthetic control method hence further confirms our previous findings.

1.4 Conclusion

This study uses newly-collected data on WWI volunteering in 121 cities in England and Wales to examine the effect of public shaming on volunteering rates. Using information from local newspapers, I build a measure for the activity of the White Feather Girls and examine how volunteering rates change after a city receives the first news about the movement. In a difference-in-difference design with staggered treatment I find that the White Feather Girls had a significantly positive impact on volunteering in the ten days after the news was received.

The results highlight a new aspect of social image concerns: even concerns evoked by strangers have the power to induce potentially fatal decisions. In particular, I find that men take extreme actions when they see their masculinity questioned. As pointed out by Virginia Woolf[27] "[e]xternal observation would suggest that a man still feels it a peculiar insult to be taunted with cowardice by a woman in much the same way that a woman feels it a peculiar insult to be taunted with unchastity by a man." The White Feather Girls tapped directly into these emotions. By publicly humiliating men in civilian clothes, young women also suggested that their chances

27. Virginia Woolf, Three Guineas (1938), https://gutenberg.net.au/ebooks02/0200931h.html.

for finding a mate would be low. While many psychologists consider shame a dysfunctional emotion at the individual level (e.g. Dickerson, Gruenewald, and Kemeny 2004), it might act as an important driver of moral and pro-social behavior, and hence be useful on the group level (Beall and Tracy 2020). Experimental evidence suggests that public shaming can successfully alter pro-social behavior in settings with low stakes such as charitable giving and tax compliance (DellaVigna, List, and Malmendier 2012; Perez-Truglia and Troiano 2018). This paper thus also provides evidence on the functionality of shame in a real-world setting with extremely high stakes and underlines the importance of social emotions in inducing altruistic behavior.[28][29]

Later in the war, the British government itself utilized women and also children to shame men into volunteering.[30] As this study shows, public shaming *can* be highly effective in inducing altruistic behavior. However, it remains to be studied under which conditions shaming achieves its desired outcome. It is also possible that some individuals react defiantly when shamed in public. This study raises the question about the morality of "nudging" and other forms of exploiting feelings as a matter of public policy, trying to enforce the law or to establish norms (e.g. Nussbaum 2009).

28. The role of shame in shaping social norms has been studied theoretically (Bénabou and Tirole 2006). Using a structural approach and experimental data, Butera et al. (2022) analyze the welfare effects of shame.

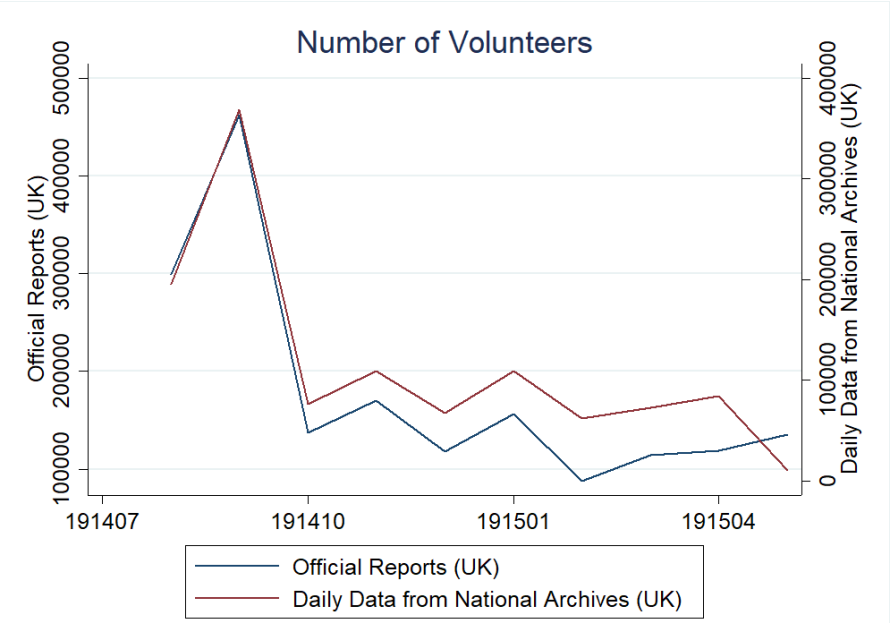
29. Recent findings suggest that shame is a universal system and part of our cooperative biology rather than a product of cultural evolution (Sznycer et al. 2018).

30. Figure 1.F.7 in the Appendix shows examples of official recruitment posters used by the government.

Appendix to Chapter 1

1.A Recruiting Data

Figure 1.A.1: Comparison of Recruiting Numbers



Note: The figure compares the recruiting numbers from official reports (left axis) to the numbers digitized from the historical records from the National Archives (Kew) (NATS 1/398) for the period between August 1915 and May 2915.

1.B Debate in the House of Commons

Figure 1.B.2: Discussion in House of Commons

WHITE FEATHERS (INSULTING CONDUCT).

[548]
43. Mr. CATHCART WASON asked the Home Secretary if he is aware that persons employed directly or indirectly in the service of the State are subjected to insolence and provocation at the hands of some advertising young women presenting them with white feathers; and if he will give orders to the police to arrest such persons for acting in a manner likely to create a breach of the peace?

Mr. McKENNA: I agree with my hon. Friend in thinking the practice very objectionable and not likely to assist recruiting, but I am informed by the Commissioner of Metropolitan Police that no complaint in the matter has reached the police, and I do not think the risk of the practice leading to breach of the peace is so great as to justify the action which my hon. Friend suggests.

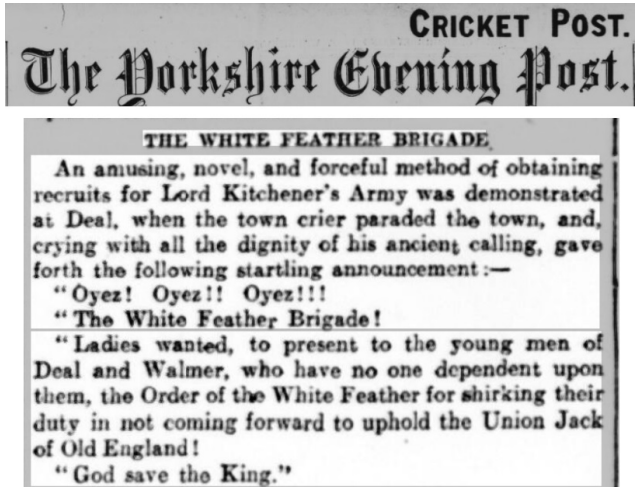
Mr. C. WASON: May I ask the right hon. Gentleman whether, if he is unable to assist in the manner indicated, he would make it easier for badges to be issued to persons employed?

Mr. McKENNA: That question should be addressed to the naval and military authorities.

Note: Excerpt from transcription of debate in the House of Commons on Monday, 1st March, 1915, column 548 ff. https://parlipapers.proquest.com/parlipapers/docview/t71.d76.cds5cv0070p0-0005.

1.C Local Newspapers

Figure 1.C.3: Newspaper Article on the White Feather Girls



CRICKET POST.

The Yorkshire Evening Post.

THE WHITE FEATHER BRIGADE.

An amusing, novel, and forceful method of obtaining recruits for Lord Kitchener's Army was demonstrated at Deal, when the town crier paraded the town, and, crying with all the dignity of his ancient calling, gave forth the following startling announcement:—

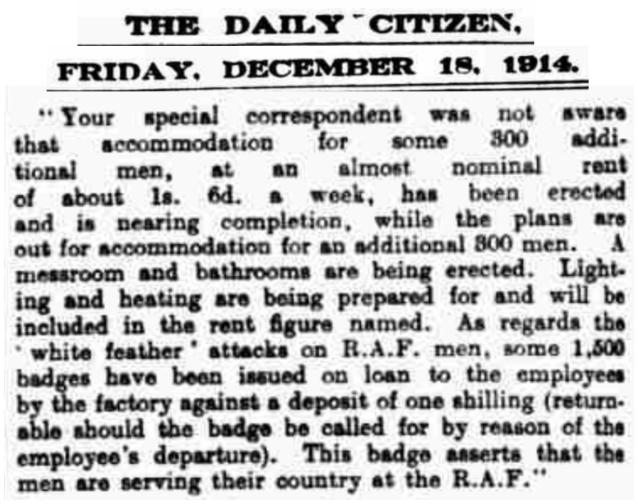
"Oyez! Oyez!! Oyez!!!

"The White Feather Brigade!

"Ladies wanted, to present to the young men of Deal and Walmer, who have no one dependent upon them, the Order of the White Feather for shirking their duty in not coming forward to uphold the Union Jack of Old England!

"God save the King."

(a) Article on White Feather Girls



THE DAILY CITIZEN,

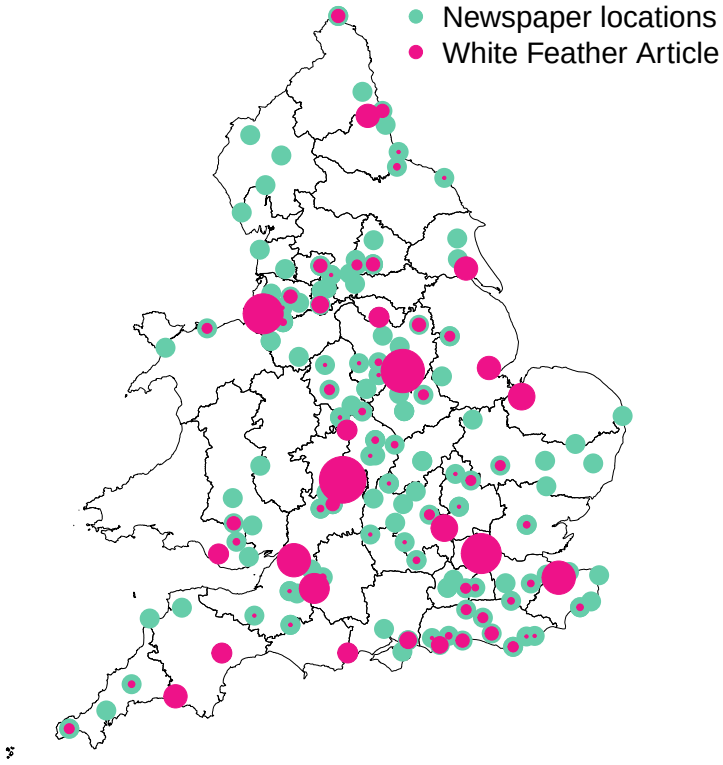
FRIDAY, DECEMBER 18, 1914.

"Your special correspondent was not aware that accommodation for some 300 additional men, at an almost nominal rent of about 1s. 6d. a week, has been erected and is nearing completion, while the plans are out for accommodation for an additional 300 men. A messroom and bathrooms are being erected. Lighting and heating are being prepared for and will be included in the rent figure named. As regards the 'white feather' attacks on R.A.F. men, some 1,500 badges have been issued on loan to the employees by the factory against a deposit of one shilling (returnable should the badge be called for by reason of the employee's departure). This badge asserts that the men are serving their country at the R.A.F."

(b) Article on Intriduction of Badges

Note: Panel (a): Example of a newspaper article on the White Feather Girls from Wednesday 02 September 1914 published in the Yorkshire Evening Post. Found in The British Newspaper Archive (britishnewspaperarchive.co.uk, subscription required). Panel (b): Example of a newspaper article discussing badges to allow men to signal employment in war-relevant industries and hence offer protection from the White Feather Girls. Published in The Daily Citizen on Friday 18 December 1914. Found in The British Newspaper Archive (britishnewspaperarchive.co.uk, subscription required).

Figure 1.C.4: Places of Publication and Number of White Feather Articles



Note: The Figure shows the locations of all 253 local newspapers in the data (green dots) and the places where articles on the White Feather Girls were published (pink dots) weighted by the number of articles published in each newspaper between August 1914 and May 1915. The maximum number of articles published on the White Feather Girls by a single newspaper was 14.

1.D Balancing

Table 1.D.1: Sample Balancing

	Full sample mean	Control mean	Treated mean	Difference
Demographics				
Age	28.69	28.86	28.56	0.293
	(1.500)	(1.191)	(1.708)	(1.07)
Share of females in population	0.51	0.51	0.52	-0.00845*
	(0.021)	(0.022)	(0.020)	(-2.21)
Share of foreigners in population	0.01	0.01	0.01	0.000988
	(0.011)	(0.013)	(0.009)	(0.49)
Share young (less than 14yo)	0.30	0.30	0.30	-0.00547
in population	(0.024)	(0.019)	(0.027)	(-1.24)
Share old (more than 65yo)	0.06	0.06	0.06	0.00195
in population	(0.015)	(0.013)	(0.016)	(0.73)
Share of working age (15 to 64yo)	0.64	0.64	0.64	0.00418
in population	(0.021)	(0.022)	(0.021)	(1.06)
Share in labor among those	0.69	0.69	0.69	0.000952
of working age	(0.037)	(0.045)	(0.030)	(0.14)
Share self-employed among those	0.08	0.08	0.08	0.000938
of working age	(0.019)	(0.019)	(0.020)	(0.26)
Share paid among those	0.45	0.44	0.46	-0.0135
of working age	(0.080)	(0.091)	(0.071)	(-0.92)
Share of farmers among those	0.27	0.30	0.24	0.0513
of working age	(0.170)	(0.163)	(0.174)	(1.66)
Share of individuals within each occupation:				
Legislators, officials, managers	0.02	0.02	0.02	0.000336
	(0.006)	(0.007)	(0.005)	(0.30)
Professionals	0.04	0.04	0.04	-0.000190
	(0.010)	(0.011)	(0.010)	(-0.10)
Technicians and ass. professionals	0.01	0.01	0.01	0.000479
	(0.007)	(0.007)	(0.006)	(0.39)
Clerks	0.05	0.05	0.05	-0.000888
	(0.023)	(0.027)	(0.020)	(-0.21)
Service workers	0.26	0.25	0.26	-0.00227
	(0.063)	(0.058)	(0.067)	(-0.20)
Skilled agricultural/fishery workers	0.15	0.16	0.13	0.0334
	(0.096)	(0.096)	(0.095)	(1.92)
Crafts and related trades workers	0.30	0.27	0.32	-0.0444*
Observations	121	54	67	121

Continued on next page.

Table 1.D.1 – continued from previous page

	Full sample mean	Control mean	Treated mean	Difference
	(0.115)	(0.105)	(0.120)	(-2.15)
Plant and machine operators	0.10	0.10	0.10	0.000954
	(0.056)	(0.054)	(0.058)	(0.09)
Elementary occupations	0.07	0.07	0.07	0.00385
	(0.020)	(0.018)	(0.021)	(1.07)
Armed forces	0.01	0.02	0.01	0.00871
	(0.026)	(0.032)	(0.020)	(1.82)
Share of women who are married	0.36	0.36	0.36	0.00383
	(0.017)	(0.013)	(0.019)	(1.25)
Share of men who are married	0.37	0.37	0.37	-0.00486
	(0.022)	(0.024)	(0.019)	(-1.22)
Share of women between 15 and 45 yo	0.52	0.52	0.51	0.00306
	(0.023)	(0.023)	(0.023)	(0.73)
Share of men between 15 and 45 yo	0.65	0.64	0.65	-0.00731
	(0.028)	(0.036)	(0.018)	(-1.45)
News coverage and sentiment				
Share of articles topic 1	0.09	0.09	0.09	-0.00489
	(0.045)	(0.045)	(0.046)	(-0.59)
Share of articles topic 2	0.17	0.16	0.17	-0.0160
	(0.069)	(0.069)	(0.069)	(-1.26)
Share of articles topic 3	0.11	0.11	0.11	0.00222
	(0.057)	(0.058)	(0.057)	(0.21)
Share of articles topic 4	0.11	0.11	0.11	0.00117
	(0.052)	(0.054)	(0.050)	(0.12)
Share of articles topic 5	0.12	0.12	0.12	0.00281
	(0.056)	(0.058)	(0.055)	(0.27)
Negative news sentiment	0.23	0.23	0.23	-0.00626
	(0.100)	(0.102)	(0.098)	(-0.34)
Observations	121	54	67	121

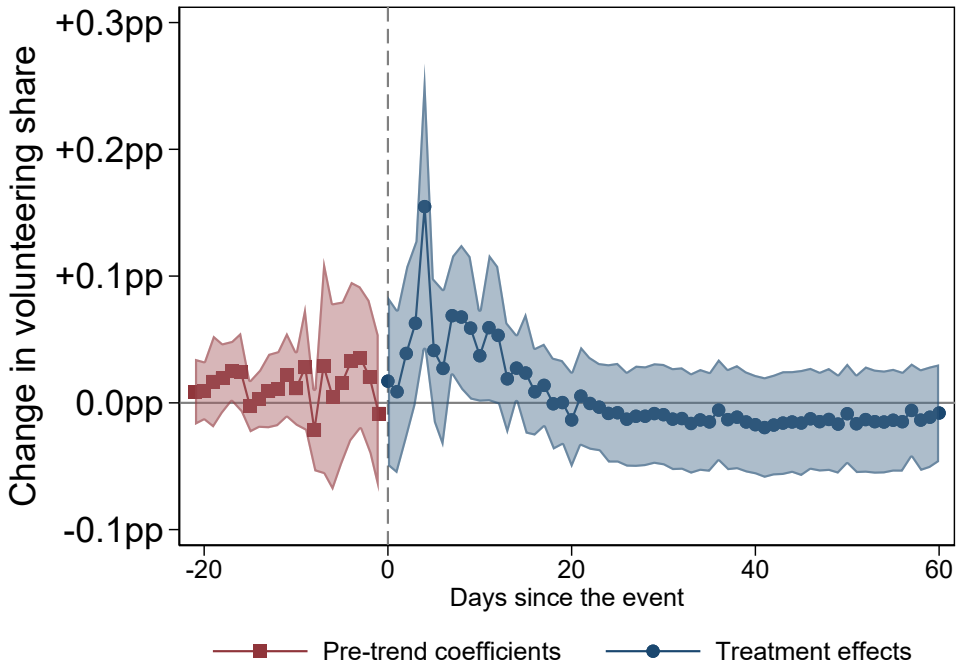
Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table shows the mean of several demographic variables taken from the 1911 Census for the whole sample and for the control group and the treated group separately as well as the difference between the control and the treated group. The census was used on the level of parishes and all parishes were assigned to the nearest recruiting city to obtain relevant demographics. Variables on news sentiment and coverage were computed as described in Section 1.3.4.2. The topics are defined as determined by the Latent Dirichlet Algorithm. For details see also Section 1.3.4.2 and in particular Figure 1.8.

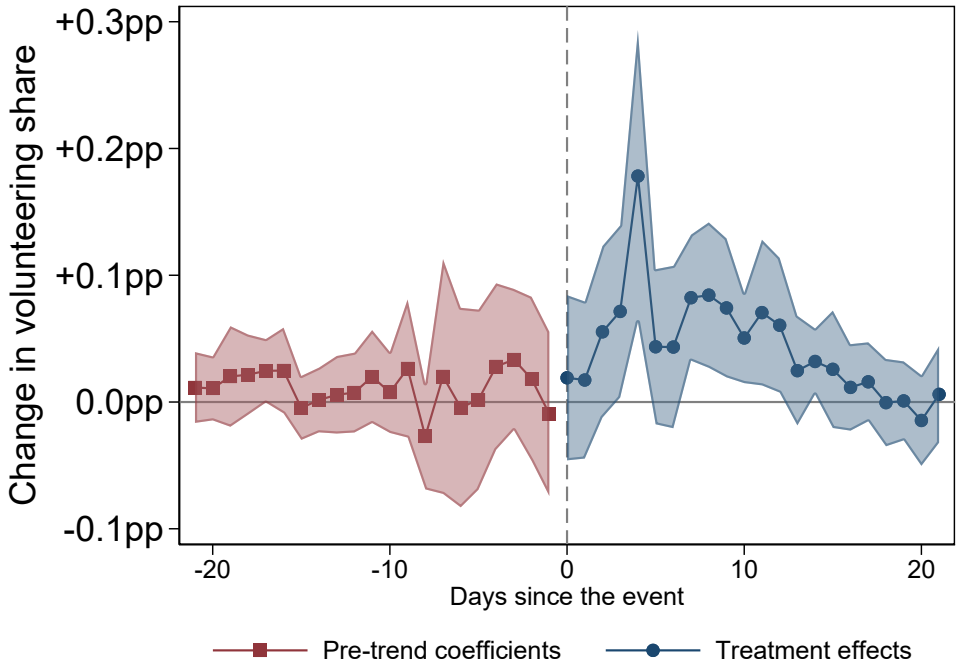
1.E Additional Results

Figure 1.E.5: Effect of White Feather Activity on Volunteering – Longer Horizon



Note: The graph shows the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample comprises all 121 recruiting cities. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of recruitable men from the 1911 Census.

Figure 1.E.6: Effect of White Feather Activity on Volunteering – Full Set of Controls



Note: The graph shows the results for pre-trend testing (red squares) based on equation (1.3) and treatment effect estimation (blue dots) using the imputation approach for difference-in-differences designs proposed by Borusyak, Jaravel, and Spiess (2021). The shaded red and blue areas indicate the 95% confidence intervals for the estimates. The sample comprises all 121 recruiting cities. Effect and pre-trends are shown for three weeks around the event date which is set to two days before the publication date of an article. A city is defined as treated after it has been closest to a newspaper publishing on the White Feathers for the first time. The outcome of interest is the volunteering share defined as the number of daily volunteers divided by the number of recruitable men from the 1911 Census. Controls include the daily share of articles within a radius of 50km around the recruiting cities falling into each of the five topics defined by the LDA as depicted in Figure (1.8). Topic 3 serves as the reference. I also control for negative sentiment measured as the daily share of articles dominantly associated with a negative emotion (anger, fear, disgust and sadness) within a 50km radius around the recruiting cities. Controls from the Census comprise the share of females, the share of the population born in another country, the share of young individuals in the population, the share of self-employed, the share of farmers and the share of those in paid employment. Casualties are the number of soldiers who died each day whose residence is in an area assigned to the respective recruiting cities. Information on soldiers who died come from the Naval & Military Press.

1.F Government Efforts using Public Shaming

Figure 1.F.7: Newspaper Article on the White Feather Girls

(a) PCR recruitment poster using female pressure

(b) PCR recruitment poster using children

Note: Panel (a): Recruiting Poster. Creator unknown. Printed by David Allen and Sons Ltd., Harrow, London. Production date: February 1915. Imperial War Museum, London. https://www.iwm.org.uk/collections/item/object/28305. Panel (b): Parliamentary Recruiting Committee, London. Poster No. 79. Designed and Printed by Johnson, Riddle and Co., Ltd., London, S.E.. Produced in 1915. Imperial War Museum, London. https://www.iwm.org.uk/collections/item/object/17053.

Chapter 2

Motivated Political Reasoning: On The Emergence of Belief-Value Constellations *(with Kai Barron and Steffen Huck)*

Abstract

We study the causal relationship between moral values ("ought" statements) and factual beliefs ("is" statements) and show that, contrary to predictions of orthodox Bayesian models, values exert an influence on beliefs. This effect is mediated by prior political leanings and, thus, contributes to increasing polarization in beliefs about facts. We study this process of motivated political reasoning in a pre-registered online experiment with a nationally representative sample of 1,500 individuals in the US. Additionally, we show that subjects do not distort their beliefs in response to financial incentives to do so, suggesting that deep values exert a stronger motivational force.

2.1 Introduction

Why would Republicans hold different beliefs about the dangers of Covid-19 than Democrats and, as a consequence, take fewer precautions and increase their risk of becoming infected? This startling pattern observed in beliefs and behavior during the pandemic is documented in both Allcott et al. (2020) and Clinton et al. (2021). However, the wider phenomenon of partisan "bubbles" comprising *disagreement about facts* along the political spectrum can be traced back in the United States to the early 2000s (see, for example, Gaines et al. 2007, on polarized beliefs about the Iraq war).

One explanation for this phenomenon is provided in a seminal article by Taber and Lodge (2006) that introduces the notion of "motivated skepticism" to explain the maintenance of partisan beliefs as a consequence of biased information processing: citizens exhibit a tendency to interrogate arguments and information that is in conflict with their prior partisan attitudes more vigorously, while uncritically accepting attitudinally congruent arguments. Furthermore, the motive to conform with one's political party (or signal one's conformity with it to others) has been shown to be an important mediator for motivated reasoning on policy issues. For example, in Druckman, Peterson, and Slothuus (2013) arguments in favor of, or against, a motion are shown to have a stronger effect on partisanship when the arguments are explicitly linked to party stances. More recently, in a conceptual piece that also reviews earlier empirical evidence, Alesina, Miano, and Stantcheva (2020) note that Democrats and Republicans appear to view the same factual reality through different lenses, which the authors refer to as a partisan "polarization of reality". Importantly, the authors note that there are several potential explanations for this phenomenon, and that the direction of causality is often unclear. For example, individuals may select

into political parties on the basis of their beliefs or other personal characteristics.

We provide a novel contribution to this line of inquiry by exploring how beliefs depend on underlying *values* – examining the correlations and causal relationships, as well as the consequences for actions. We conceptualize values as desires about the social world, that is, statements about how the world ought to be. In contrast to beliefs, which pertain to states of the physical world, values can neither be objectively true nor false – they can only be endorsed or opposed with possibly different strength.

For a rational decision maker, values might causally depend on beliefs but not vice versa. For example, if I believe that animals suffer from pain in a similar fashion to humans, I might desire a world in which animals are no longer killed and eaten.[1] In contrast, values should (according to the rational model) have no impact on beliefs. Whether I desire a world in which we are all vegetarians should not influence my assessment of how likely it is that animals suffer from pain when they are injured.

Yet, if values are central to an individual's identity, such that maintaining them is important, then motivated reasoning might be employed in protecting these core values. Such motivated reasoning could reverse the causal relationship, allowing values to causally shape the beliefs individuals hold.

We study this relationship between values and beliefs in the context of six different contentious policy domains: migration, animal welfare, gender equality, abortion, prostitution and same-sex marriage. In doing so, we reveal the role played by party preferences in motivated reasoning and also analyze the consequences for actions,

1. Notice that this maneuver does not contradict Hume's (1978) assessment that ought statements cannot be derived purely from a set of facts. Support for the value statement that animals should not be killed requires implicitly drawing on an ought axiom about the avoidance of suffering in general.

specifically, donations to charities operating in these six domains.

While values have been shown to be strongly associated with party preferences and voting decisions (see, for example, Enke 2020a) the precise causal relationship between values and factual beliefs remains largely unexplored. In our study we can draw on previously measured party preferences and show how they mediate the relationship between values and beliefs.[2]

In studying the role played by values in the (motivated) formation of beliefs, we contribute to a growing literature on motivated cognition and wishful thinking. This literature has considered a wide array of factors that may generate motivated beliefs, including: maintaining a positive image of one's own intelligence or attractiveness (e.g., Eil and Rao 2011; Möbius et al. 0; Coutts 2019; Drobner 2021), judging what is fair or morally appropriate in a self-interested fashion (e.g., Messick and Sentis 1979; Babcock et al. 1995; Konow 2000; Barron, Stüber, and Veldhuizen 2019; Amasino, Pace, and Van der Weele 2021), distorting one's own beliefs in order to be more persuasive to others (e.g., Schwardmann and Van der Weele 2019; Solda et al. 2020; Schwardmann, Tripodi, and Van der Weele 2021), and engaging in confirmatory reasoning that reinforces one's prior beliefs (e.g., Nickerson 1998; Rabin and Schrag 1999).

Previous literature typically considers motivated reasoning in a relation to a belief

2. To the extent that values may drive party preferences, one could capture the full causal system that we have in mind as a directed acyclic graph (DAG) in the spirit of Pearl (2009) with beliefs being (potentially) driven by values and party preferences, and party preferences also influenced by values (i.e., $v \rightarrow p \rightarrow b \leftarrow v$). In the language of DAGs party preferences would then be a mediator between values and beliefs, and values occupy the parent node, influencing beliefs either directly or via party preferences. Note, also, that any rational model would require that causality run in the opposite direction, with beliefs serving as a parent node to values and party preferences.

that is closely tied to an individual's self-interest, their personal characteristics or their pre-existing beliefs about a particular topic. In contrast, here we examine whether deeper values may exert an influence over related factual beliefs. Given the extent to which many contentious political debates are driven by values and given also the substantial heterogeneity in values between and within societies, this strikes us as an important question. Individuals who differ in their values might still be able to achieve compromises as long as they agree on the facts. But when the facts are in dispute and beliefs about the very nature of the world diverge, bipartisan action will be severely impeded.

In order to explore the role played by moral values in influencing factual beliefs, as well as the joint influence of beliefs and values on decision making, we designed and conducted a pre-registered online experiment in January 2020 that surveyed a nationally representative sample of 1,500 individuals from the US population.[3] Our experiment comprises four main treatments that allow us to test hypotheses organized around the following questions: (i) whether there exist systematic correlations between values and factual beliefs; (ii) whether individuals adapt their factual beliefs when there is an increase in the salience of a moral value in the same domain; (iii) whether individuals shift their stated values and factual beliefs in order to align them with their own material self interest; and (iv) whether individuals adjust their stated values and beliefs in an attempt to persuade others.

To measure beliefs about factual statements, we ask subjects "How likely do you think

3. To provide some context, the experiment was, thus, designed and implemented prior to events such as the widespread awareness of COVID-19 (February 2020), the death of George Floyd (May 2020), the claims that the United States presidential election was rigged (November 2020) and the attack on the Capitol (January 2021).

it is that the following statement is true?"; for moral statements we ask "How much do you agree with the following statement?" This reflects that for facts there is, in principle, an ascertainable truth while values can only be desirable or undesirable to different degrees.

To fix ideas, let us consider two examples of statement pairs, the first from the migration domain, the second from the animal welfare domain. The statement "All countries benefit from the free movement of labor" pertains to a fact that may either be true or false. While its veracity might be difficult to ascertain, it is, in principle, ascertainable. In contrast, the statement "People should be allowed to migrate freely between countries" expresses a desire. One may or may not agree with the statement but there is no truth to be ascertained. This is similarly the case for the two statements "Animals feel less pain than humans" (belief) and "It is wrong to eat animals" (value).

Our study makes use of these two pairs of statements as well as four others created in a similar way. That is, in each case the factual statement and the moral statement come from the same domain of life such that we can employ the notion of belief-value constellations that can be spatially represented as "bubbles" or constellations in two-dimensional belief-value space.[4]

Our first two treatments address questions (i) and (ii) – whether there are correlations between values and beliefs and whether values do indeed causally shape beliefs. First, in treatment ValueNotSalient we elicit subjective beliefs about factual

4. Essentially, the notion of belief-value constellations reflects the idea that the beliefs and values individuals hold might manifest in clusters of associated beliefs and values, with individuals who hold a certain value more likely to hold a certain belief, and vice versa.

statements from the six domains mentioned above. This treatment serves as a control for the second treatment, ValueSalient, where we additionally elicit subjects' agreement with value statements pertaining to the six domains prior to the belief elicitation. This elicitation of values serves to raise the salience of these values when subjects report their associated beliefs thereafter. Of course, subjects may also be passively aware of their values in treatment ValueNotSalient but their direct elicitation in ValueSalient should serve as a priming device that heightens value salience, bringing them to the forefront of the individual's mind. The underlying idea for treatments ValueSalient and ValueNotSalient is that when faced with a heightened salience of the value question, individuals may shift their factual beliefs in a motivated way in order to align them with their values. We test this by comparing the distribution of beliefs reported in these two treatments. We examine this comparison both unconditional and conditional on party preferences.

In all our treatments there is one final stage after the belief and value elicitations exploring choices that relate to the six domains. Specifically, we give subjects the opportunity to donate money to charities that operate in each of the six domains. This final stage plays a key role for our next two treatments that address questions (iii) and (iv) – whether subjects are willing to distort their values and beliefs in order to convince themselves or to convince others. To test whether self-interest plays a role in shifting beliefs, in treatment ConvinceSelf, we put the donation decision on the same screen on which we elicit beliefs and values (in other treatments, it comes as a surprise). We can, hence, test whether subjects adjust their reasoning in a self-interested way when holding particular belief-value constellations would point towards taking a costly action (i.e., making a donation to a charity whose work is

aligned with that particular constellation). In such a setting, individuals might be less inclined to hold certain beliefs and/or values as it becomes more costly to do so. Finally, we test whether subjects adjust their stated beliefs and/or values when they have an incentive to persuade others. In the fourth treatment, ConvinceOther, we again ask subjects to state their beliefs and values but rather than making the donation decision themselves as in the third treatment, they are informed that another participant will have the opportunity to donate after being shown their belief and value responses.

We find strong support for the existence of aligned belief-value constellations in all the policy domains considered, thereby answering our first research question in the affirmative and providing crisp evidence for the existence of "bubbles" with within-bubble homogeneity of beliefs and values and between-bubble heterogeneity in both dimensions. Notice, however, that while such correlations are evidence for some kind of partisanship it is not possible to understand the mechanism behind this without further evidence. Such bubbles may arise when beliefs shape values (in a fully rational manner), when there are filter bubbles or echo chambers (Flaxman, Goel, and Rao 2016; Enke 2020b), or when values provide a sufficiently strong force for motivated reasoning. It is the comparison of our first two treatments that helps to explore the latter mechanism.

In the aggregate, the distributions of reported subjective beliefs are almost identical across the two treatments which, on the surface, appears to suggest that we do not observe a shift in subjects' beliefs when making values more salient. However, the picture changes dramatically when we control for individuals' political preferences. Specifically, we find that subjects on both the political right and the political left shift

their beliefs to align them with the average beliefs held by those in their preferred political party when made aware of the associated value debate. We, therefore, do find support for the idea that values shape beliefs through motivated reasoning. As a consequence, heightened salience of contentious policy issues in public debate emerges as an explanatory force for the increasing polarization in factual beliefs along political attitude division lines. This is a new result that is subtly different from motivated skepticism or other forms of biased updating previously documented in the political science literature and adds a new dimension to the motivated reasoning literature in economics—an individual's deep values can motivate their reasoning about factual beliefs as easily as monetary incentives or self-image does in other contexts.

We also find the effect of values to be quite large. The magnitude of the shift in beliefs due to value salience is nearly as large as the baseline difference in beliefs between subjects on the left and the right in the control treatment. We conduct several robustness exercises in support of these results, including replacing our main measure of political attitude with various alternative proxy measures that were elicited by us, as well as measures elicited independently by the recruitment platform prior to our experiment.

Our work therefore contributes empirical evidence to an active recent theoretical discussion about how and why partisan individuals increasingly seem to live with polarized mental models of reality (Leeper and Slothuus 2014; Van Bavel and Pereira 2018; Alesina, Miano, and Stantcheva 2020). In particular, we relate closely to Bonomi, Gennaioli, and Tabellini (2021), who discuss a theory of identity politics where increasing the salience of a certain policy conflict leads individuals to identify more strongly with their cultural or economic group, and then to distort their beliefs

towards the stereotypical belief of the group they identify with. Therefore, the main results of our paper can be viewed as providing support for some of the central ideas expressed in their theory (although our experiment was not designed as a test of their theory).[5]

With respect to our third and fourth research questions, we find beliefs and values unaffected by the addition of monetary incentives to persuade oneself or the anticipation of the opportunity to persuade another person. If anything, this lack of malleability of beliefs and values to other factors appears to suggest that our subjects care about responding honestly to our belief and value questions which should lend credibility to the internal and external validity of our first two sets of results. These results are also consistent with a growing body of research documenting the limits of motivated reasoning. In particular, several of the studies that examine whether belief updating is distorted by monetary incentives associated with different states of the world fail to find any influence of motivated reasoning (see, for example, Gotthard-Real 2017; Coutts 2019; Barron 2021). Furthermore, Thaler (2020) convincingly shows an absence of positivity-motivated reasoning in domains where self-image is not present. Specifically, the author shows that people don't engage in motivated reasoning in forming beliefs about whether the world is a good place for others to live

5. While we focus predominantly on assessing the causal effect of values on beliefs, we also contribute to a broader literature that examines the influence of partisanship on information processing. For example, in the domains of energy policy and climate change respectively, Bolsen, Druckman, and Cook (2014) and Druckman and McGrath (2019) examine the role played by partisan differences in information processing due to selectively trusting different *sources* of information. Kahan (2013) explores the role of different thinking styles in generating ideological polarization and Alesina, Stantcheva, and Teso (2018) show that when individuals are provided with pessimistic information about mobility, left-wing individuals become more pessimistic about mobility and increase their demand for redistribution, but right-wing individuals do not. In our paper, individuals are not provided with any new information to process—they must form their beliefs based on the information already stored in their memory. We only vary the presence of a reason for motivated reasoning, such as the salience of a policy conflict.

(i.e., about cancer survival rates, infant mortality and others' happiness). Together, these results indicate that motivated reasoning operates in certain domains, with internal psychological factors such as self-image and deep values serving as a source for motivated reasoning, but external factors such as monetary rewards and others' well-being often not resulting in motivated reasoning.

2.2 Existence and Formation of Belief-Value Constellations

Our experimental design consists of four pre-registered[6] treatments that were conducted online using the platform Prolific with a nationally representative sample of 1,863 individuals from the US population.[7] In this section we focus on describing and analyzing the first two treatments, ValueSalient and ValueNotSalient, which allow us to ask: *(i) Do individuals display belief-value constellations? (in the sense of observing a correlation between beliefs and values), and (ii) Do individuals adjust their beliefs to be more coherent with their values?*

2.2.1 The ValueSalient and ValueNotSalient treatment conditions

The objective of our ValueSalient treatment is to examine whether we observe a systematic correlation between values and associated factual beliefs. The experimental design of the ValueSalient treatment consists of three parts. First, participants are presented with a sequence of six (randomly ordered) moral value statements and

6. The full pre-registration document can be found at https://osf.io/8jydh/ and is also reproduced in Appendix 2.C. This research was approved by the UCL Research Ethics Committee (REC 17181/001).

7. Table 2.B.1 in the appendix shows that our sample is strongly balanced between all the treatments which are described in the following.

are asked to report the degree to which they agree or disagree with the statement. Each of these moral value statements corresponds to a particular contentious topic of debate in public policy, such as gender equality, abortion or same-sex marriage. Therefore, the moral value agreement questions serve to raise the salience of these debates for the participants, who might then view later questions in the experiment through the lens of those debates. Second, participants are asked to state their belief that each of the six factual statements is true. Importantly, each of the six factual statements is related to one of the same six public policy debates as the moral value statements.[8] Together, the moral value assessments and factual belief reports allow us to examine whether there is a correlation between individuals' beliefs and values. Third, to examine how these values and beliefs translate into actions, we provide participants with the opportunity to make six donation decisions to six separate charities (one of which is randomly implemented). Each of the six charities targets a cause that corresponds to of one of the six relevant public policy discussions.[9] For each charity, participants are asked to divide $3 between the charity and themselves. In a post-experimental survey, we also collected information on the participants' political attitudes and, additionally, we were able to match our data to previously elicited political attitude variables collected by Prolific independently from our experiment.

The ValueNotSalient treatment is identical to the ValueSalient treatment, with the exception that the first stage in which participants are presented with moral

8. The moral value statements are evaluated on a 5-point Likert scale from "Strongly Agree" to "Strongly Disagree", while the factual beliefs are also assessed on a 5-point Likert scale from "Very Unlikely" to "Very Likely". The public policy debates that we consider are migration, animal welfare, gender equality, abortion, prostitution and gay rights. A complete overview of the moral value and factual statements can be found in Table 2.C.1 in Appendix 2.C.

9. Subjects are provided with information about the aims of the charities and use a slider to indicate how much they would like to donate. Further details about the charities can be found in Table 2.C.2 in Appendix 2.C.

value statements is skipped. This implies that in this treatment the six public policy debates are not made as salient. The exogenous variation in salience between the two treatments allows us to assess how this shift in salience affects the factual beliefs.

2.2.2 The Existence of Belief-Value Constellations

The first question we seek to answer with these two treatments is whether there is an alignment between the moral values, factual beliefs and political attitudes that individuals hold. This would indicate the presence of "belief-value constellations". It is important to note that such belief-value constellations are not a natural implication of standard economic theory, where individuals process information and update their beliefs about factual statements in a dispassionate way according to Bayes' rule. However, there are several potential reasons why individuals might form aligned beliefs and values, including: i) the *avoidance of cognitive dissonance* from holding incoherent values and beliefs, and ii) the use of value and belief statements to justify self-interested actions (i.e., *motivated reasoning*).

The presence of such belief-value constellations would imply that it is important to take an individual's moral values into consideration when trying to understand belief formation regarding factual statements (which is not typically done in the literature). Our first set of hypotheses address this question of whether belief-value constellations are observed systematically in the population.[10]

10. In the interest of facilitating a more coherent exposition of the paper and to enhance readability, we have adjusted the formulation of the hypotheses in comparison to the pre-registration document. We encourage the interested reader to refer to the full pre-registration document in Appendix 2.C for further details.

Hypothesis 1: Belief-Value Constellations

There is a correlation between the beliefs, values, and political attitudes that individuals hold. The actions individuals take are aligned with their belief-value constellations.

Let b_t denote the factual beliefs stated by individuals in Treatment $t \in \{VS, VNS\}$, v_t the moral values stated by individuals, d_t their donation decisions and p_t the left-right political stance of individuals.

a) Moral values are positively correlated with beliefs:

$$Corr(v_{VS}, b_{VS}) \geq 0.$$

b) Moral values are negatively correlated with political attitudes:

$$Corr(v_{VS}, p_{VS}) \leq 0.$$

c) Donations are positively correlated with beliefs and values:

$$Corr(d_{VNS}, b_{VNS}) \geq 0,$$

$$Corr(d_{VS}, b_{VS}) \geq 0, Corr(d_{VS}, v_{VS}) \geq 0.$$

When reading Hypothesis 1, it is important to take note of the way that the variables are encoded. First, the political stance variables, p_t, are constructed to be increasing in the degree to which an individual positions herself on the right of the political spectrum. Second, the moral value, v_t, variables are encoded such that a high value

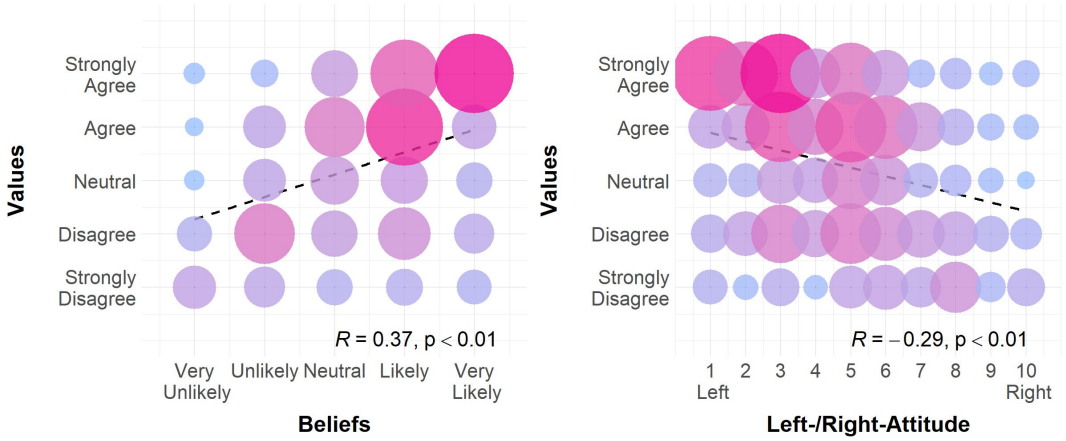
indicates agreement with a value that is typically associated with individuals on the left of the political spectrum. Third, the factual belief variables, b_t, are defined such that if they are true, they would provide empirical support for moral value positions typically held by individuals on the political left. Finally, the charitable donation variables, d_t, are constructed such that higher donations are consistent with costly support of a charity aligned with the relevant moral value position.

Results (Hypothesis 1)

In this section, we provide evidence relating to the key hypothesis of whether individuals form beliefs and values in a manner that generates belief-value constellations. Figure 2.1 summarizes the results pertaining to this hypothesis. First, the top left panel reports the correlation between beliefs and values across all policy domains. This shows a strong positive relationship between beliefs and values that is statistically significant at the 1% level. Second, the top right panel shows the results for the correlation between values and political attitudes. In line with the hypothesis, we observe a negative relationship, with left-leaning political attitudes associated with higher agreement with the moral value statements. Third, the three panels in Figure 2.2 show that donations are positively correlated with beliefs in the ValueNotSalient treatment, and are also positively correlated with beliefs and values in the ValueSalient treatment. All three are statistically significant at the 1% level.

Collectively, these results are completely in line with the pre-registered set of hypotheses, providing strong evidence for the presence of belief-value constellations. This suggests that individuals form beliefs, values and political attitudes in a manner that generates strong associations between the different objects. This is important

Figure 2.1: Results for Hypothesis 1a and 1b



(a) Beliefs and Values (Hyp. 1a) **(b)** Political Attitudes and Values (Hyp. 1b)

Note: Figure 2.1(a) shows the results on Hypothesis 1a, i.e. the correlation between values and beliefs in the ValueSalient treatment. Figure 2.1(b) shows the results for Hypothesis 1b, i.e. the correlation between moral values and political attitudes in the ValueSalient treatment. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of values on beliefs respectively political attitudes. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

and striking because economists often conceptualize *values* as being preferences held by individuals, while *beliefs* reflect probability assignments over states of the world and pertain to the (objective) processing of information. These objects are typically treated as being orthogonal by economists. Our results suggest a more nuanced interdependent relationship between these two objects.

In Appendix 2.A.1, we reproduce these results for each of the six topics separately. While most of the topic-specific results are completely in line with the aggregate analysis above, the analysis reveals several interesting findings regarding the heterogeneity in the strength of the associations between the variables across topics. While Figure 2.A.1 shows that all six topics display a positive relationship between beliefs and values at the 1% level, Figure 2.A.2 shows that there is a negative relationship be-

tween values and political attitudes for five of the six topics. The prostitution topic in fact displays a statistically significant *positive* relationship between political attitudes and values, with individuals who identify with the political right stating stronger agreement that that prostitution should be illegal than individuals on the political left.

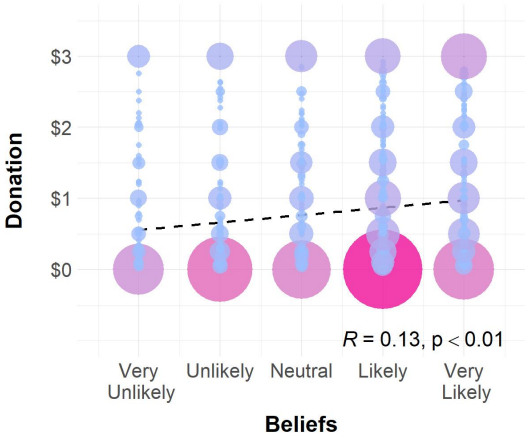
Figures 2.A.3, 2.A.4 and 2.A.5 illustrate the relationship between donation decisions and beliefs and values in the ValueNotSalient and ValueSalient treatments. While the point estimates of these relationships are positive in all eighteen comparisons, we observe substantial heterogeneity in the strength of the relationship. Overall, the relationship between donations and both values and beliefs appears to be weakest for the prostitution-related charity, which received relatively high donation levels across all beliefs and values. Interestingly, for the abortion-related charity, the relationship was very weak in the ValueNotSalient treatment, but very strong when the value debate was made salient in the ValueSalient treatment. This is indicative of a shift in contribution choices when seen through the lens of the contentious value debate.

Overall, the results show strong support for Hypothesis 1.

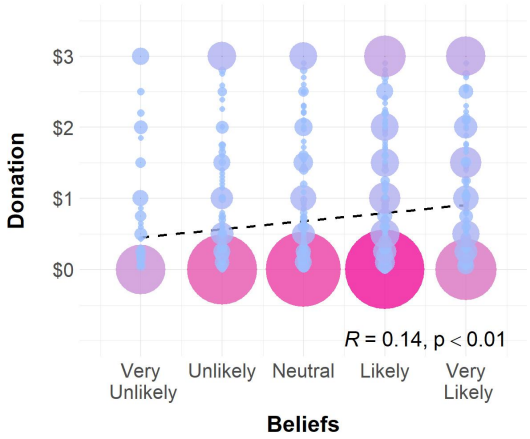
2.2.3 The Formation of Belief-Value Constellations

Our second hypothesis asks whether the formation of factual beliefs is influenced by the salience of a particular contentious moral value debate. *Do individuals adjust their factual beliefs when examining them with a related hotly contested moral issue at the forefront of their mind?* If this is the case, it would speak to the question of why individuals end up holding tightly clustered beliefs and values.

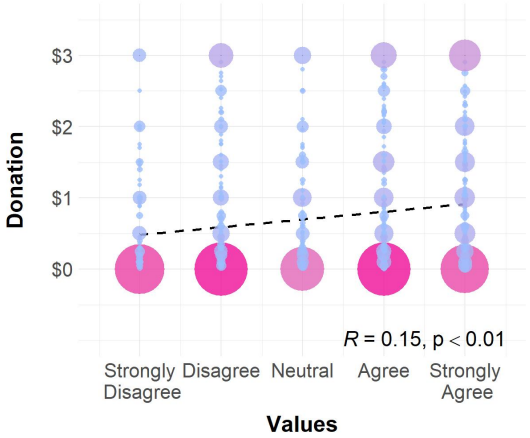
Figure 2.2: Results for Hypothesis 1c



(a) Donations and Beliefs in VNS



(b) Donations and Beliefs in VS



(c) Donations and Values in VS

Note: The three figures show the results on Hypothesis 1c. Figure 2.2(a) shows the correlation between beliefs and donations in treatment VALUENOTSALIENT, Figure 2.2(b) shows the correlation between beliefs and donations in treatment VALUENOTSALIENT, and Figure 2.2(c) shows the correlation between values and donations in treatment VALUESALIENT. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

To test this, we use a between-treatment comparison of the distribution of beliefs observed in the VALUENOTSALIENT and VALUESALIENT conditions. We can thus assess whether factual beliefs are shifted when we prime individuals to think about these

belief statements through the lens of the related value debate.

This is formalized in Hypothesis 2 below, which posits that: (i) the salience of values affects belief formation, and (ii) this mechanism can result in the polarization of factual beliefs. The rationale for this is that if (i) is true then the heterogeneity in moral values between different political groups would also lead to the formation of these polarized factual beliefs. This would provide one potential explanation for observed recent trend of increasingly polarized factual beliefs along ideological lines (see, e.g., Gentzkow 2016; Enke 2020a) which has been documented in various domains such as *climate change* (McCright and Dunlap 2011) and *COVID-19 beliefs* (Allcott et al. 2020).[11]

Hypothesis 2: Construction of Consistent Beliefs

Increasing the salience of a contentious moral value debate leads individuals to report factual beliefs that are more strongly aligned with their moral value. This results in an increase in polarization of factual beliefs.

Let F_{b_t} denote the cumulative distribution function (cdf) of factual beliefs b_t in treatment $t \in \{VS, VNS\}$, F_{v_t} the cdf of moral values v_t, and F_{d_t} the cdf of donations d_t. As before, p_t denotes the left-right political stance of individuals.

a) Raising the salience of a moral value influences factual beliefs.

The distribution of factual beliefs differs between the ValueNotSalient and Value-

11. In his theoretical work, Le Yaouanq (2021) links heterogeneity in political attitudes to partisan disagreement about objective facts through people's idiosyncratic preferences regarding the policy implications of scientific findings. Our work seeks to understand the underlying psychological mechanisms in more detail.

SALIENT treatments:

$$F_{b_{VNS}} \neq F_{b_{VS}}.$$

b) *A higher degree of polarization in values in a particular policy domain will result in a stronger effect of increased value salience on the dispersion of factual beliefs.*

Comparing across the six domains indexed by m, the difference between the variance in beliefs in VALUESALIENT and the variance in beliefs in VALUENOTSALIENT is non-decreasing in the variance in values in VALUESALIENT:

$$\frac{d[Var(b^m_{VS}) - Var(b^m_{VNS})]}{d[Var(v^m_{VS})]} \geq 0.$$

c) *Raising the salience of a moral value results in an increase in polarization of factual beliefs.*

Beliefs in VALUESALIENT are more polarized than beliefs in VALUENOTSALIENT:

$$E(b_{VS}|p_{VS} < E(p_{VS})) - E(b_{VNS}|p_{VNS} < E(p_{VNS}))$$
$$\geq$$
$$E(b_{VS}|p_{VS} > E(p_{VS})) - E(b_{VNS}|p_{VNS} > E(p_{VNS})).$$

Several features of this set of hypotheses are worth highlighting. First, the rationale for part b) and c) of the hypothesis is that the raised salience of the relevant value will result in a shift towards more extreme factual beliefs as subjects are drawn towards more coherent belief-value constructions. Second, the inequality in part c) states that the difference between the average factual belief in VALUESALIENT versus VALUENOT-

SALIENT is greater for subjects on the left of the political spectrum in comparison to those on the right. To put this another way, the hypothesis states that individuals on the left will increase their beliefs between VALUENOTSALIENT and VALUESALIENT more than individuals on the right of the political spectrum, on average.[12]

RESULTS (HYPOTHESIS 2)

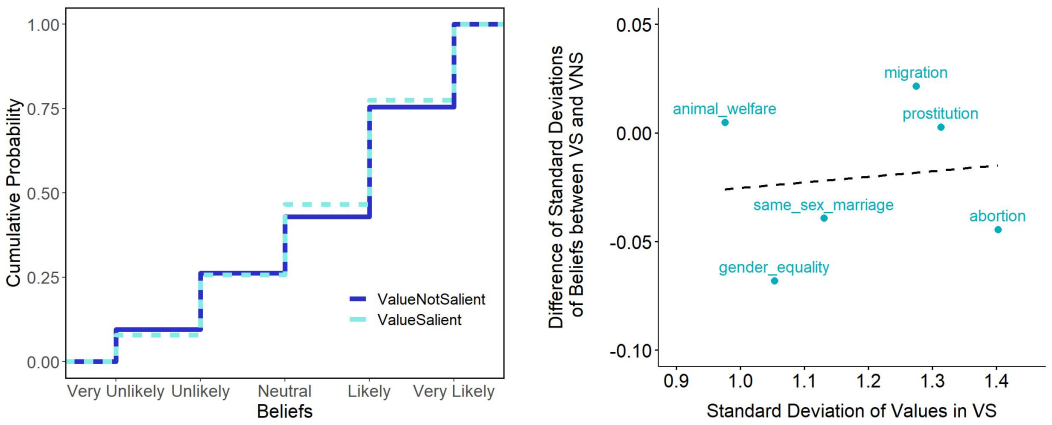
In this section we examine the relationship between values and beliefs more closely by asking whether raising the salience of a particular value leads to a causal shift in an associated factual belief. This comparison represents a fairly conservative test of the existence of a causal relationship between beliefs and values for several reasons. First, our experiment focuses on short-run motivated reasoning and does not consider causal effects of motivated cognition that operate over a longer period of time (e.g., via biased information search or selective memory). Second, our experiment exploits a salience manipulation of values, which represents a fairly weak dosage of the treatment (i.e., an exogenous shift of values). We therefore view our treatment manipulation as placing a lower bound on the causal relationship between values and beliefs.

Figure 2.3 provides a summary of the results associated with Hypothesis 2a and 2b. The left panel displays the cumulative distribution of reported beliefs in the VALUENOTSALIENT and VALUESALIENT treatments. This panel provides suggestive evidence in favor of the polarization of beliefs when pooling across all topics, with the dashed line weakly below the solid line in the left part of the figure, while the dashed line is weakly above the solid line in the right part of the figure. However, this

12. Note that this also includes the case where subjects on the right adjust their beliefs downwards which would make the right-hand side negative.

difference is not statistically significant at the 5% level (p-value = 0.09, Kolmogorov-Smirnoff test). Second, the right panel of the figure asks whether there is a heterogeneous effect of increasing the salience of a particular topic. For topics with a high degree of variance in values (i.e. highly polarized issues), we hypothesized that increasing the salience of these values would lead to a larger degree of polarization of the VALUESALIENT beliefs relative to the beliefs in VALUENOTSALIENT (Hypothesis 2b). While we do observe an upward sloping linear relationship, the slope coefficient is not statistically different from 0.

Figure 2.3: Results for Hypotheses 2a and 2b



(a) CDF of Beliefs and Values in VS and VNS

(b) Change in Variance of Beliefs as a Function of Value Spread

Note: Figure 2.3(a) shows the results for Hypothesis 2a, i.e. the cumulative density function of beliefs in treatments VALUESALIENT and VALUENOTSALIENT. Figure 2.3(b) shows the result for Hypothesis 2b. The y-axis shows the difference of the standard deviations of beliefs between treatments VALUESALIENT and VALUENOTSALIENT and the x-axis shows the standard deviation of values in treatment VALUESALIENT. The dotted line depicts the result from a linear regression of the difference of the standard deviations on the standard deviation of values.

The results for Hypotheses 2a and 2b suggest that raising the salience of values did not result in a clear shift in the distributions of beliefs across all six issues. Hypothesis 2c posits that even if an increase in value salience does not result in an increase in polarization of the aggregate distribution of beliefs, there may be heterogeneity in

the impact of the value salience at the individual level—i.e., the political preferences of an individual could mediate how increasing the salience of their values shifts their beliefs. Essentially, Hypothesis 2c asserts that making a value more salient leads individuals to shift their beliefs even further towards conforming with the average beliefs held by members of their own political party.

To address this question, we therefore compare the belief movement of individuals on the left of the political attitude spectrum with those on the right of the political attitude spectrum. Using a difference-in-difference style empirical approach, we ask whether the gap between the beliefs of those on the left and the right increases in the ValueSalient treatment relative to in the ValueNotSalient treatment. To do this, we estimate the following regression:

$$b_{i,j} = \alpha_1 \cdot \tilde{p}_{i,j} + \alpha_2 \cdot ValSal_{i,j} + \alpha_3 \cdot \tilde{p}_{i,j} \times ValSal_{i,j} + \epsilon_{i,j} \tag{2.1}$$

where $b_{i,j}$ is the reported belief of individual i for topic j, $ValSal_{i,j}$ is a binary variable that equals 1 if the individual is in the ValueSalient treatment and 0 when the individual is in the ValueNotSalient treatment, and $\tilde{p}_{i,j}$ is an indicator variable that takes a value of 1 if the individual is on the left of the political spectrum (i.e. reports a political attitude that is lower than the mean political attitude reported in our sample).

The coefficient of interest is α_3, corresponding to the interaction term. This essentially compares how individuals on the left and right of the political spectrum change their factual beliefs when exposed to an increase in value salience. A positive coef-

ficient denotes a widening of the gap between the factual beliefs of the left and the right. Table 2.1 reports the results from estimating equation 2.1 in the first column, with VALUESALIENT × Pol. Attitude denoting the interaction term. The estimates show that the coefficient on the interaction term is positive and statistically significant at the one percent level, providing evidence that we do indeed observe polarization of the beliefs along political attitude division lines when related contentious values are made salient. It is worth noting that this increase in polarization is on top of the pre-existing difference in factual beliefs reported between individuals on the left and right of the political spectrum in the VALUENOTSALIENT treatment. This is shown by the significant coefficient associated with the Pol. Attitude variable. It is also worth noting that the size of the widening of the gap in factual beliefs between the left and right due to the salience is nearly as large as the baseline difference in factual beliefs between individuals on the left and the right in VALUENOTSALIENT (i.e. the magnitude of the coefficient associated with the variable *VALUESALIENT×Pol. Attitude* is $\frac{3}{4}$ the size of the coefficient associated with the variable *Pol. Attitude*).

In order to test the robustness of this result, we conduct several additional exercises. First, we check whether the results are driven by the specific political attitudes variable that we have chosen to use. To do this, we run two further regression, where we replicate the estimation in the first column of Table 2.1, but replace the *Left* indicator variable with a variable that indicates that the individual self-reported being a *Democrat* (Column (2)) and a variable that indicates that the individual voted for Hilary Clinton in 2016 (Column (3)). The results from both of these exercises are highly consistent with our main estimation results in Column (1).[13]

13. Importantly, both of these variables were collected by Prolific completely separately from our experimental data collection. Therefore, these results also serve to alleviate possible concerns regard-

Table 2.1: Influence of increased salience of values on belief polarization

	(1)	(2)	(3)
ValueSalient	-0.124**	-0.199**	-0.236***
	[0.052]	[0.087]	[0.075]
Pol. Attitude ($\tilde{p}$)	0.269***	0.358***	0.304***
	[0.047]	[0.071]	[0.058]
ValueSalient	**0.199*****	**0.294*****	**0.315*****
× Pol. Attitude	**[0.067]**	**[0.100]**	**[0.089]**
Constant	3.325***	3.202***	3.264***
	[0.038]	[0.063]	[0.049]
Observations	4560	2550	3006
Pol. Attitude ($\tilde{p}$) Variable	Left-Right Scale (Left = 1)	Party Affiliation (Democrat = 1)	Last Election (Clinton = 1)

Clustered standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: ValueSalient is a dummy variable equal to one if the individual was assigned to treatment ValueSalient and hence equal to zero if the individual was assigned to treatment ValueNotSalient. We use three measures of the political attitudes variable. This is indicated in the last two rows of the table. In column (1), Political Attitude is a dummy equal to one if the individual is below the median on a 1 to 10 scale of political attitudes where 1 is the most left and 10 is the most right attitude. In column (2), Political Attitude is a dummy equal to one if the individual identifies as a Democrat rather than as a Republican and in column (3) Political Attitude equals one if the individual indicated that they voted for Clinton in the 2016 elections and zero if they voted for Trump.

Second, instead of collapsing the 10-point political attitudes variable into a binary indicator (as in Table 2.1), we estimate this regression using the full 10-point scale and report all ten interaction coefficients in Figure 2.4. This shows that individuals on the left of the political attitude spectrum are shifting their beliefs upwards relative to individuals who reported being completely on the right (i.e., those who reported a 10 and constitute the omitted category). Overall these robustness exercises are

ing our political attitudes variable being influenced by the treatment condition. However, a caveat to this is that the Prolific variables are only available for a subset of the sample. This is the reason for the differing sample sizes across the three regressions.

highly supportive of the finding that when an issue is viewed as more politically charged (e.g., when a contentious related value discussion becomes more salient), individuals tend to shift their beliefs to conform more with their political in-group.

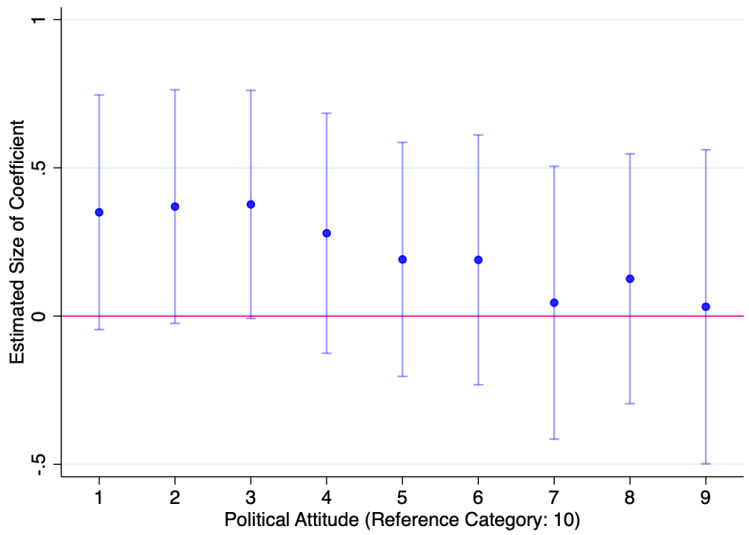
Interestingly, this is not associated with a polarization of factual beliefs at the aggregate level.[14] These results highlight an important distinction between two forms of polarization, namely (i) polarization of the entire unconditional distribution, which involves movement towards extreme beliefs, and (ii) polarization conditional on a particular characteristic (e.g., political party) that defines groups in the population. The latter form of polarization involves a reshuffling of the belief distribution and may or may not lead to aggregate or unconditional polarization. Drawing this distinction between *unconditional polarization* and *conditional polarization* is important as it helps us to understand the mechanisms in play. *Unconditional polarization* can be driven by a variety of mechanisms, such as confirmation bias or other individual cognitive heuristics that favor coherent beliefs and values, while *conditional polarization* points towards social conformity with one's in-group as a driving factor.

Overall, the collective evidence provided by the exercises and robustness checks we implement is supportive of Hypothesis 2c (i.e., *conditional polarization*), but we do not observe strong evidence in favor of Hypotheses 2a and 2b (i.e., *unconditional polarization*).[15]

14. This can be the case if the individuals from the two parties are moving their beliefs in opposite directions and essentially replacing one another out in the aggregate distribution. This occurs when the movers do not start off close to their party-aligned factual belief extreme, and therefore have to "jump over" one another to conform with their political in-group's beliefs.

15. In Appendix Section 2.A.2.1, we also document the donation behavior in the ValueSalient and ValueNotSalient treatment conditions. In summary, we do not observe evidence of a substantial effect on donation decisions, suggesting that the shift in beliefs is not translating into a change in behavior on this dimension. This result contributes to the growing body of work documenting a complex relationship between beliefs and actions.

Figure 2.4: Influence of Increased Salience on Beliefs across the Political Spectrum (Hyp. 2c)



Note: Figure 2.4 shows the coefficients α_3 estimated using a regression as the one described in equation (2.1) above with the difference that $\tilde{p}_{i,j}$ is using the full scale of political attitudes, i.e. we include a dummy for each value on the political left-right scale, using the category '10' (the most right category) as a reference. The vertical lines represent the 95% confidence interval for the estimated coefficients.

2.3 Convincing Yourself and Convincing Others

After having explored how beliefs react to values we now ask whether money exerts a similar influence on beliefs and possibly on values. The third set of our hypotheses below will be divided into two parts, with both parts assessing the malleability of beliefs and values to monetary forces that could pull them in different directions. In Part A (Convincing Yourself), we examine the role of self-serving biases in the context of belief-value constellations by asking whether individuals try to justify selfish behavior by adjusting their beliefs and values to be consistent with taking actions that are in their material self-interest—engaging in a form of motivated reasoning or excuse-driven behavior.[16] In addition, Part B (Convincing Others) studies whether

16. Previous work has shown that people develop self-serving biases in order to excuse their selfishness in charitable giving (see e.g. Exley (2015) on the role of risk or Exley (2020) on using charity performance metrics as an excuse).

introducing the opportunity to try to convince another participant to take a specific action can lead to a shift in one's own beliefs. Specifically, we ask whether attempts to engage in persuasion lead to a shift in beliefs.

2.3.1 The ConvinceSelf and ConvinceOther treatments

We conduct two further treatments. First, the ConvinceSelf treatment speaks to the conjecture that individuals adjust their beliefs and values in a self-serving way. This treatment is very similar to ValueSalient, with only one key difference: In ConvinceSelf, subjects are aware that they will need to make a charitable donation decision when they form and report their moral value and factual belief assessments. This is in contrast to ValueSalient, where the charitable donation screen arrive as a surprise after the moral value and factual belief reports have been completed. This difference is important, since subjects' anticipation of the costly charitable donation decision could influence their introspection in forming a personal assessment of the value and factual belief statements. Hypothesis 3a conjectures that individuals bias their (stated) beliefs and values when they take into account the costs of an expected donation decision, with the bias shifting beliefs and values away from those that would justify a higher donation.[17]

Second, the ConvinceOther treatment examines how trying to convince others to take an action that is in line with one's own values could lead an individual to further align their factual beliefs with their political agenda or goals, and perhaps to exaggerate these stated beliefs. To do this, treatment ConvinceOther mirrors again

17. It is also possible that the anticipated donation might operate in the opposite direction inflating the the importance of values and beliefs to convince such that the agent can convince herself that making a high donation is the correct decision. In our pre-registration document we noted this possibility but stated that our prior was that the self-serving bias would dominate.

treatment ValueSalient with just a single exception: before stating their values and beliefs, subjects are informed that *another* participant will have the option to donate to a related charity after being informed about the moral values and factual beliefs that they (the subject in ConvinceOther) reported. So, participants might reflect on the possibility that their own values and beliefs could exert and influence on the donation decision of another subject. In order to avoid deception we implemented these decisions by others in an auxiliary treatment, BeingConvinced. The first part of the BeingConvinced treatment is identical to ValueSalient, with subjects reporting their values and beliefs on the six relevant topics. The difference arrives prior to subjects making their donation decisions. At this point, subjects in BeingConvinced are informed about the beliefs and values stated by a randomly chosen participant from ConvinceOther.

2.3.2 Do individuals self-servingly shift their beliefs and values?

To examine this question, we compare behavior in ConvinceSelf, where subjects anticipate their future donation decisions, with behavior in ValueSalient, where subjects report their values and beliefs before they are aware of the future donation decisions. This allows us to study the robustness of elicited beliefs and values in the presence of monetary incentives that could distort them. We, specifically, ask whether the presence of the donation decision on the same screen induces subjects to distance themselves from the charity-aligned value position, and similarly adjust their beliefs away from supporting the charity's goals. This is summarized in the following set of hypotheses.

Hypothesis 3A: Convincing Yourself

As before, let F_{b_t} denote the cumulative distribution function (cdf) of factual beliefs b in Treatment t, and F_{v_t} the cdf of moral values v. Donations in Treatment $t \in \{VS, VNS, CS, CO\}$ are denoted by d_t and p_t denotes the left-right political stance of individuals.

a) Individuals shift their beliefs and values to justify taking self-serving actions: In ConvinceSelf individuals shift their beliefs and values downwards in comparison to in ValueSalient in order to justify low future donation decisions. Specifically:

i) b_{VS} first-order stochastically dominates b_{CS}, i.e. $F_{b_{VS}} \leq F_{b_{CS}}$.

ii) v_{VS} first-order stochastically dominates v_{CS}, i.e. $F_{v_{VS}} \leq F_{v_{CS}}$.

b) Donations in ConvinceSelf are lower than in ValueSalient:

$$E(d_2) \geq E(d_3).$$

Results (Hypothesis 3A)

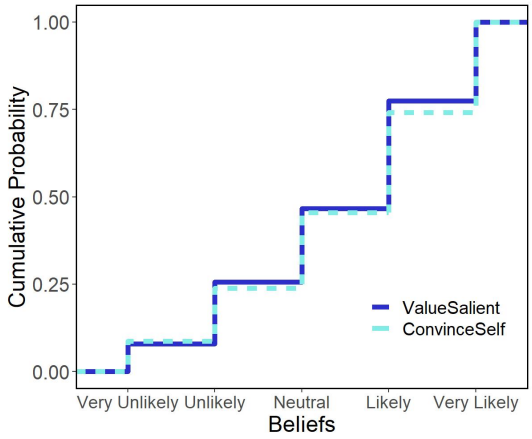
Essentially, we find no evidence in support of Hypothesis 3A. Figure 2.5 displays the distribution of beliefs (top left panel), values (top right panel) and donations (bottom panel) in the ValueSalient and ConvinceSelf treatments. We observe no significant differences in behavior between these two treatments, indicating that individuals do not shift their beliefs and values when faced with an imminent donation decision. This immutability of behavior to the anticipated donation decision is in stark contrast to the effects salient values documented above. While subjects

are engaging in politically motivated reasoning they do not engage in economically motivated reasoning. Perhaps one reason for this is that individuals place a higher value on their personal identity, which incorporates their beliefs and values, than they do on a small monetary gain that they would obtain by reducing their donation. A second factor worth noting is that a large fraction of subjects donated less than 1 dollar. Thus, the cognitive dissonance costs of donating a low amount are perhaps not sufficiently high to warrant a shift in beliefs or values to justify it.

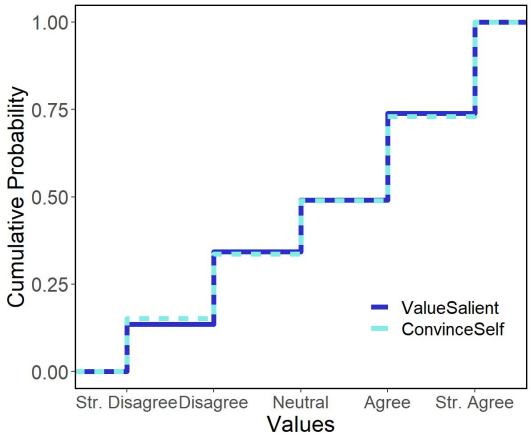
2.3.3 Do individuals shift their beliefs and values to convince others?

The second part of Hypothesis 3 asks whether individuals report more polarized factual beliefs when they have the opportunity to try to persuade someone else about the importance of certain value positions. It therefore contributes to the body of existing work that examines the idea that we adjust our own beliefs and attitudes (i.e., convince ourselves) in order to convince others Babcock et al. (1995), Schwardmann and Van der Weele (2019), Solda et al. (2020), and Schwardmann, Tripodi, and Van der Weele (2021). While this previous work predominantly studies scenarios in which an individual is explicitly mandated to convince others about a particular policy position or that they are of high ability, a key difference in our study is that we focus on examining whether individuals try to persuade others to take an action that is aligned with their own values by stating more extreme beliefs. For example, we ask whether an individual might increase their agreement with the statement that "Animals feel less pain than humans." in order to encourage another person to donate to an animal protection charity.

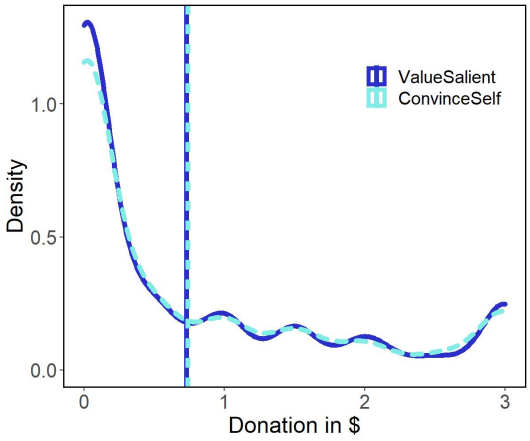
Figure 2.5: Results for Hypothesis 3A



(a) CDF of beliefs in VS and CS (Hypothesis 3A.a.i)



(b) CDF of values in VS and CS (Hypothesis 3A.a.ii)



(c) PDF of donations in VS and CS (Hypothesis 3A.b)

Note: The three figures show the results on Hypothesis 3. Figure 2.5(a) shows the cumulative density function of beliefs for treatment VALUESALIENT (dark line) and CONVINCESELF (light dotted line), Figure 2.5(b) shows the cumulative density function of values for treatment VALUESALIENT (dark line) and CONVINCESELF (light dotted line), and Figure 2.5(c) probability density function of donations for treatment VALUESALIENT (dark line) and CONVINCESELF (light dotted line). The vertical lines in Figure 2.5(c) depict the mean of donations in the two treatments.

HYPOTHESIS 3B: CONVINCING OTHERS

Anticipating the opportunity to persuade another individual about a contentious moral

issue shifts one's own factual beliefs towards the in-group party aligned extreme—i.e., factual beliefs in ConvinceOther *are more polarized than factual beliefs in* ValueSalient*:*

$$E(b_{CO}|p_{CO} < E(p_{CO})) - E(b_{VS}|p_{VS} < E(p_{VS}))$$
$$\geq$$
$$E(b_{CO}|p_{CO} > E(p_{CO})) - E(b_{VS}|p_{VS} > E(p_{VS})).$$

Similarly to Hypothesis 2c above, the inequality here in Hypothesis 3c states that the gap in factual beliefs between individuals on the left and the right of the political spectrum widens when there is an anticipated persuasion opportunity.

Results (Hypothesis 3B)

To examine Hypothesis 3c, Table 2.2 uses the same empirical specification as above and tests for a divergence of beliefs according to political attitudes between the ValueSalient and ConvinceOther treatment. Essentially, this asks whether individuals shift their beliefs even further towards conforming with their political in-group when they know that their reports will be viewed by others. The results do not support this hypothesis, with the estimated coefficient on the interaction term close to zero. Plausible explanations include: (i) that individuals do not wish to persuade others, (ii) that individuals are not prepared to adjust their own beliefs to persuade others, and (iii) that they do not believe that others will be easily persuaded in the context of these contentious debates.[18]

18. Another potential reason for this is that we are comparing the beliefs in the ConvinceOther with the ValueSalient, where beliefs have already been shifted towards political conformity relative to ValueNotSalient due to the salience of the value debates. This salience shift may crowd out any further shift when individuals wish to convince others.

Table 2.2: Regression Results (Outcome: Beliefs)

	(1)
ConvinceOther	0.089
	[0.058]
Pol. Attitude ($\tilde{p}$)	0.468***
	[0.048]
ConvinceOther × Pol. Attitude	0.005
	[0.074]
Constant	3.201***
	[0.036]
Observations	4488

Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: ConvinceOther is a dummy variable equal to one if the individual was assigned to treatment ConvinceOther and hence equal to zero if the individual was assigned to treatment ValueSalient. Political Attitude is a dummy equal to one if the individual is below the median on a 1 to 10 scale of political attitudes where 1 is the most left and 10 is the most right attitude. The outcome of interest is individuals' beliefs.

2.4 Conclusion

This paper studies the relationship between moral values and factual beliefs based on the results of a pre-registered online experiment that surveyed a nationally representative sample of 1,500 individuals from the US population. First, we ask whether there exist systematic correlations between moral values ("ought" statements) and factual beliefs ("is" statements). This is answered in the affirmative and is consistent with previous research that discusses the societal shift towards an increasingly partisan view of the world (see, e.g., Alesina, Miano, and Stantcheva 2020; Bonomi,

Gennaioli, and Tabellini 2021). As we discuss above, there are many mechanisms that might create such a correlation. For example, beliefs might shape values. Our study examines whether there exists a (reverse) causal relationship between values and beliefs where values exert an influence on beliefs which should not happen in a perfectly rational Bayesian world. We explore this by introducing a treatment that makes a moral value more salient prior to eliciting beliefs. Strikingly, while there appears to be no effect in the aggregate, a closer inspection shows substantial causal effects of values on beliefs—effects that are mediated by prior political leanings. In other words, we find that individuals in our representative sample are engaged in politically motivated reasoning.[19]

Politically motivated reasoning takes place on both sides of the political spectrum: subjects on both the political right and the political left, shift their beliefs to align them with the average party beliefs when values are made salient. This finding contrasts with the popular belief that the flirtation with "alternative facts" is a phenomenon exclusive to populist right-wing movements.

Second, we examine whether there is also evidence for economically motivated reasoning whereby individuals bias their beliefs and/or values due to the presence of monetary incentives to do so. This is not the case. We believe that this result enhances the credibility of our main findings. Since beliefs and values do not react to (small) monetary incentives, it appears that individuals care about them to the extent

19. The behavior observed in our study is consistent with the findings of Bordalo, Tabellini, and Yang (2021), who study the effect of issue salience on beliefs about others' political attitudes. The authors show that when the salience of a particular policy conflict is raised, this increases the perception of the partisan gap in attitudes. Combined with an identity-induced desire to conform to the stereotypical beliefs of one's identity group (as in Bonomi, Gennaioli, and Tabellini 2021), this perceived increase in the partisan gap could contribute to the shift in beliefs that we observe.

that they do not shade them through economically motivated reasoning.

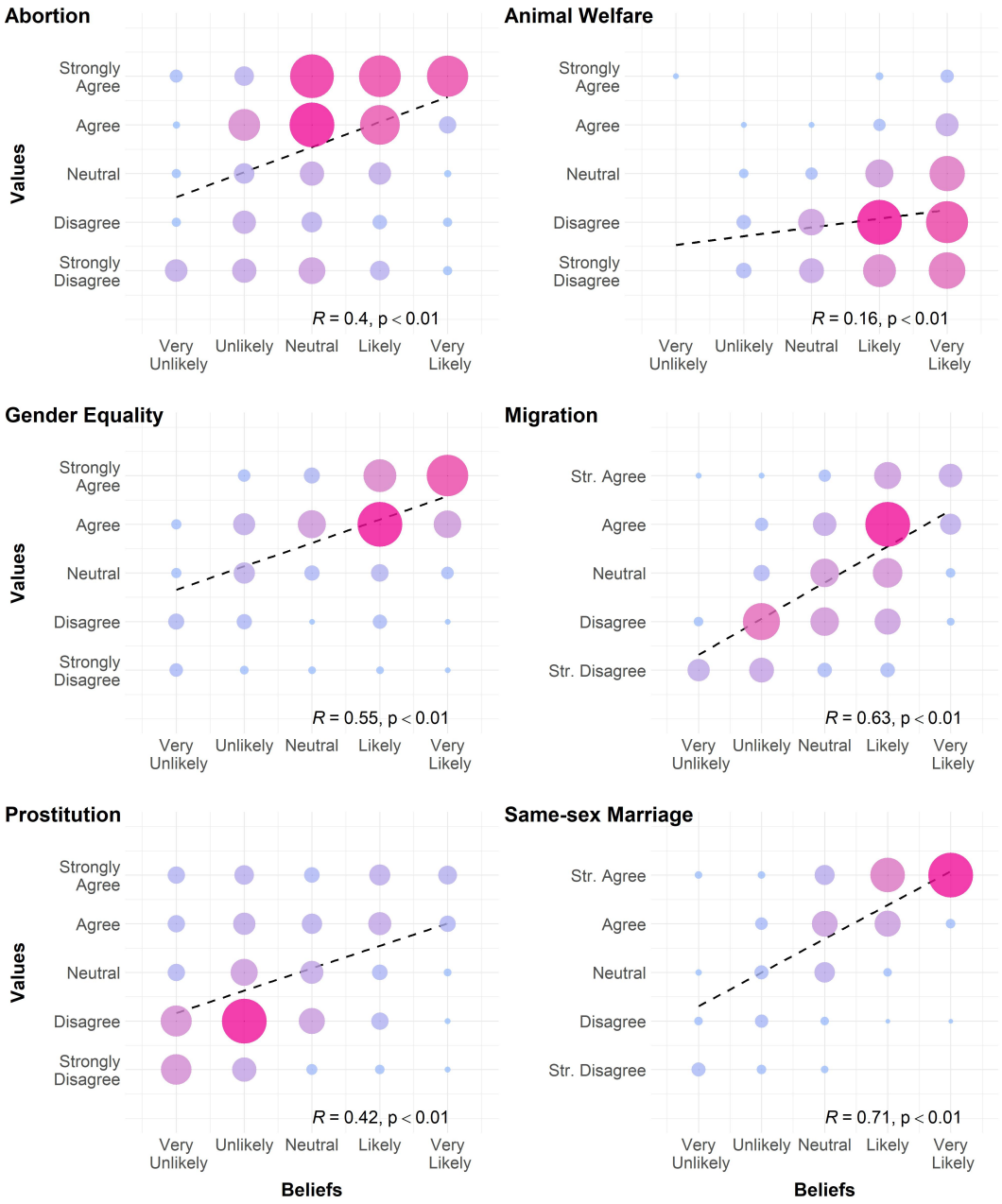
Taken together, our results point towards a deep (cognitive) link between values and beliefs. This is in sharp contrast to the notion that they should be treated as disjoint separate objects as described by the standard model. This tight relationship between values and beliefs is consistent with the conceptual idea of a "polarized reality", where individuals perceive reality through the lens of their economic or social identity (Alesina, Miano, and Stantcheva 2020) and then adjust their beliefs to conform to the stereotypical belief of the salient identity group (Bonomi, Gennaioli, and Tabellini 2021). More broadly, this recent line of research showing how identity shapes beliefs through the desire for group-conformity builds on a longer history of research examining how identity can generate a desire for conformity in *actions* (George A Akerlof and Rachel E Kranton 2000a, 2005; Shayo 2020). With the polarization of social discourse (particularly online) seemingly increasing in society, this body of work points towards identity-induced belief conformity as an important avenue for further research.

Appendix to Chapter 2

2.A Supplementary Results

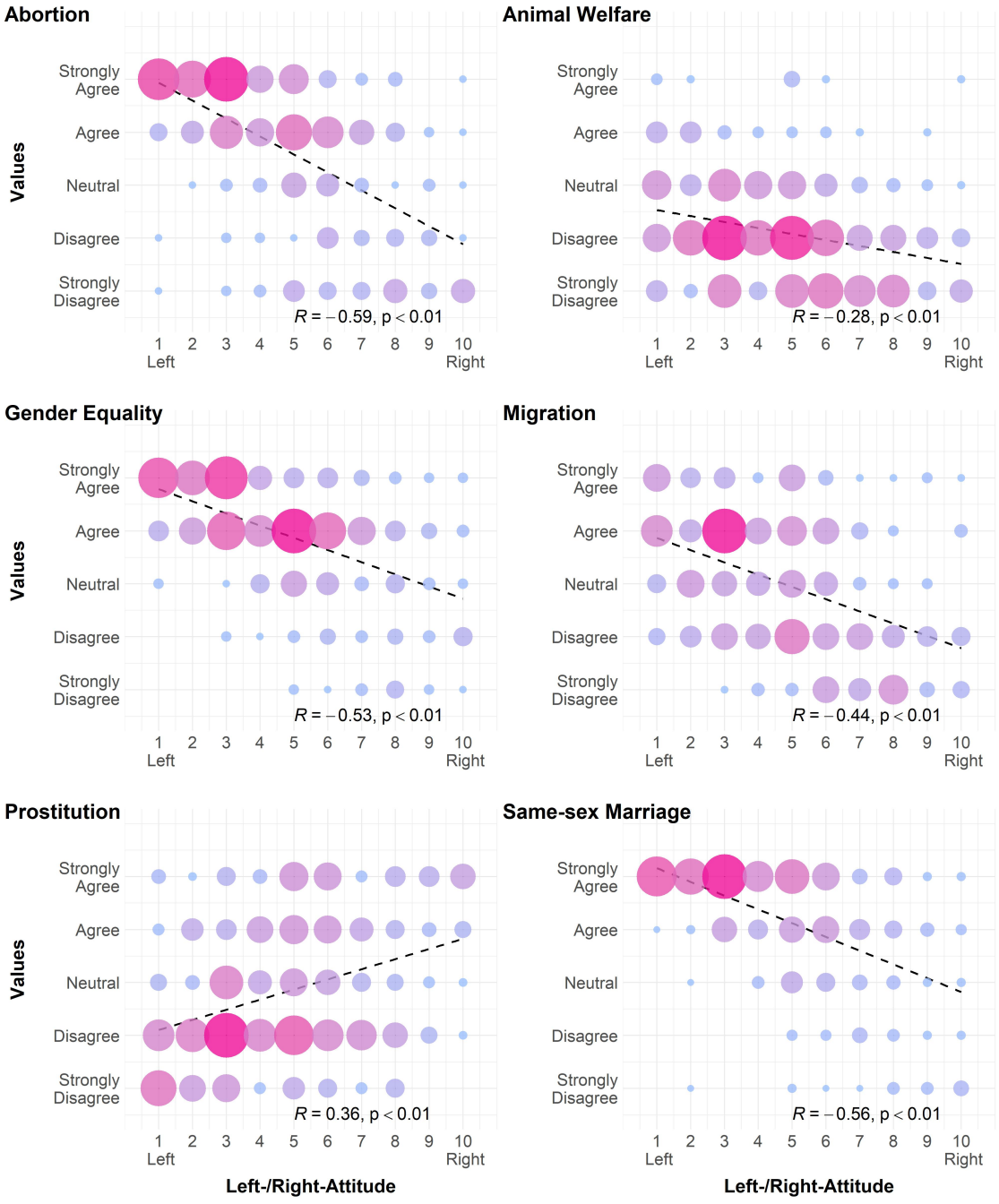
2.A.1 Supplementary Results for Hypothesis 1

Figure 2.A.1: Relationship between Values and Beliefs in ValueSalient, by Topic



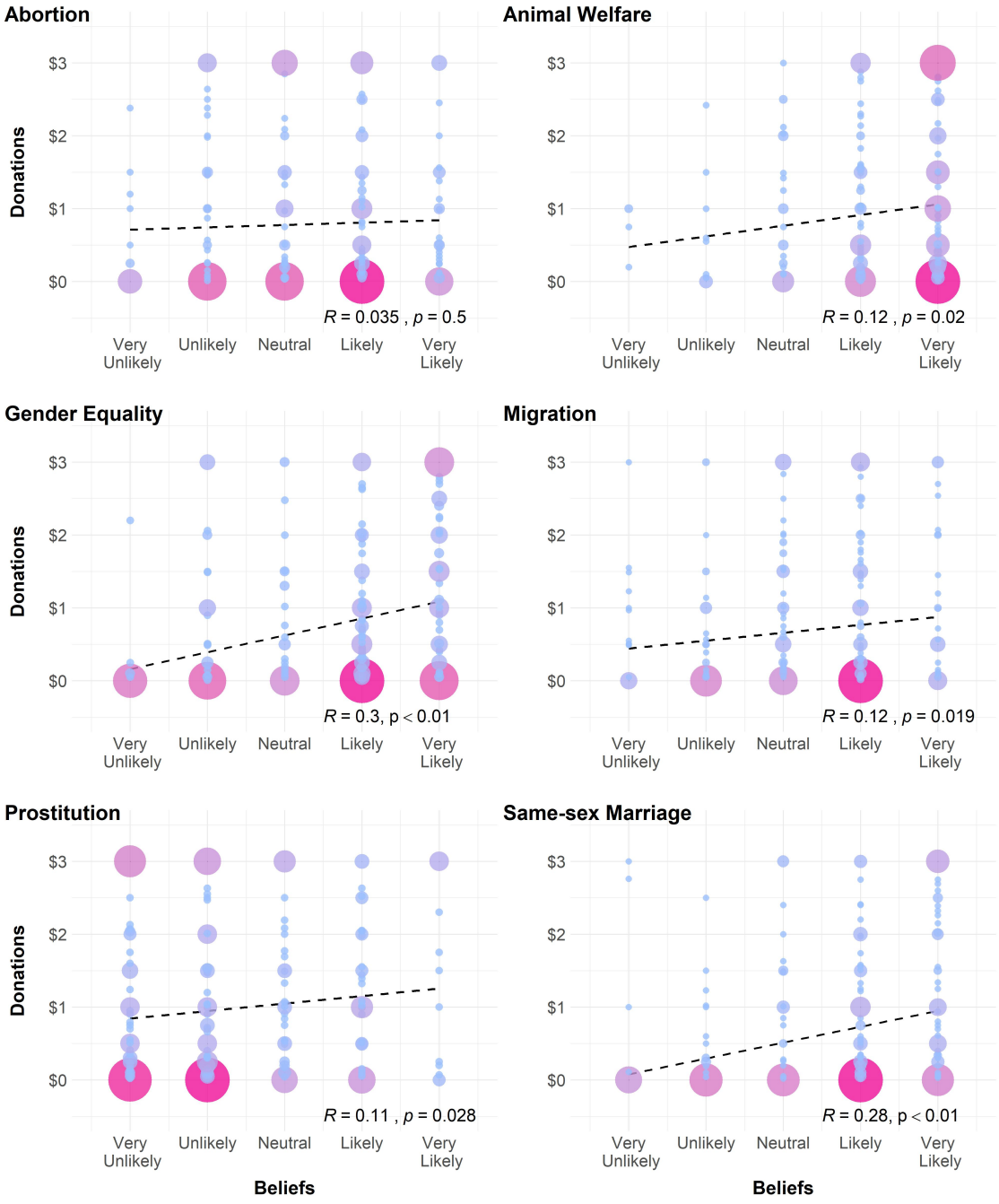
Note: Figure 2.A.1 shows the correlation between values and beliefs in the ValueSalient treatment, separately for each of the six policy debates. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of values on beliefs respectively political attitudes. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

Figure 2.A.2: Relationship between Political Attitudes and Beliefs in ValueSalient, by Topic



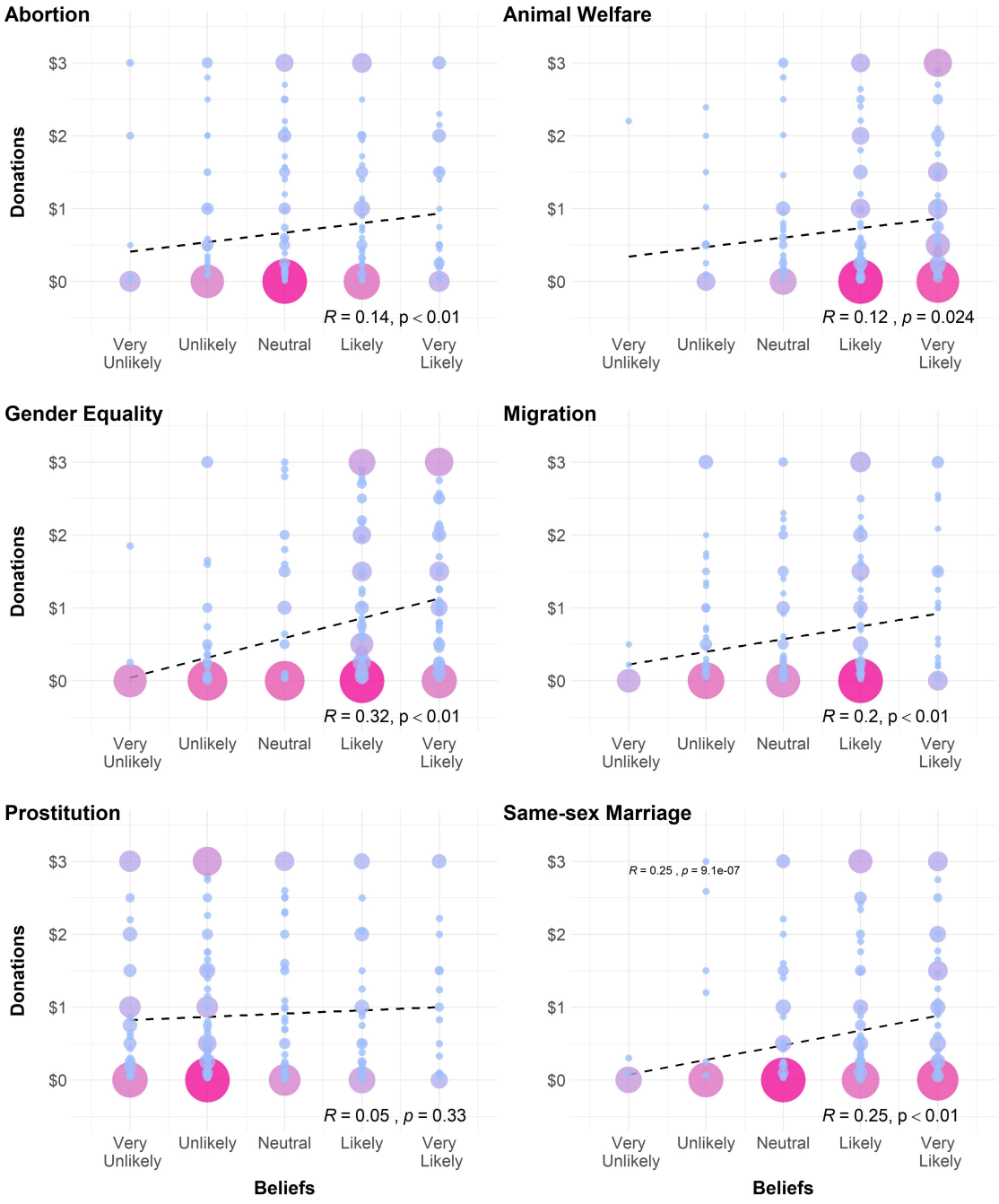
Note: Figure 2.A.2 shows the correlation between moral values and political attitudes in the ValueSalient treatment, separately for each of the six policy debates. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of values on beliefs respectively political attitudes. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

Figure 2.A.3: Relationship between Donations and Beliefs in VALUENOTSALIENT, by Topic



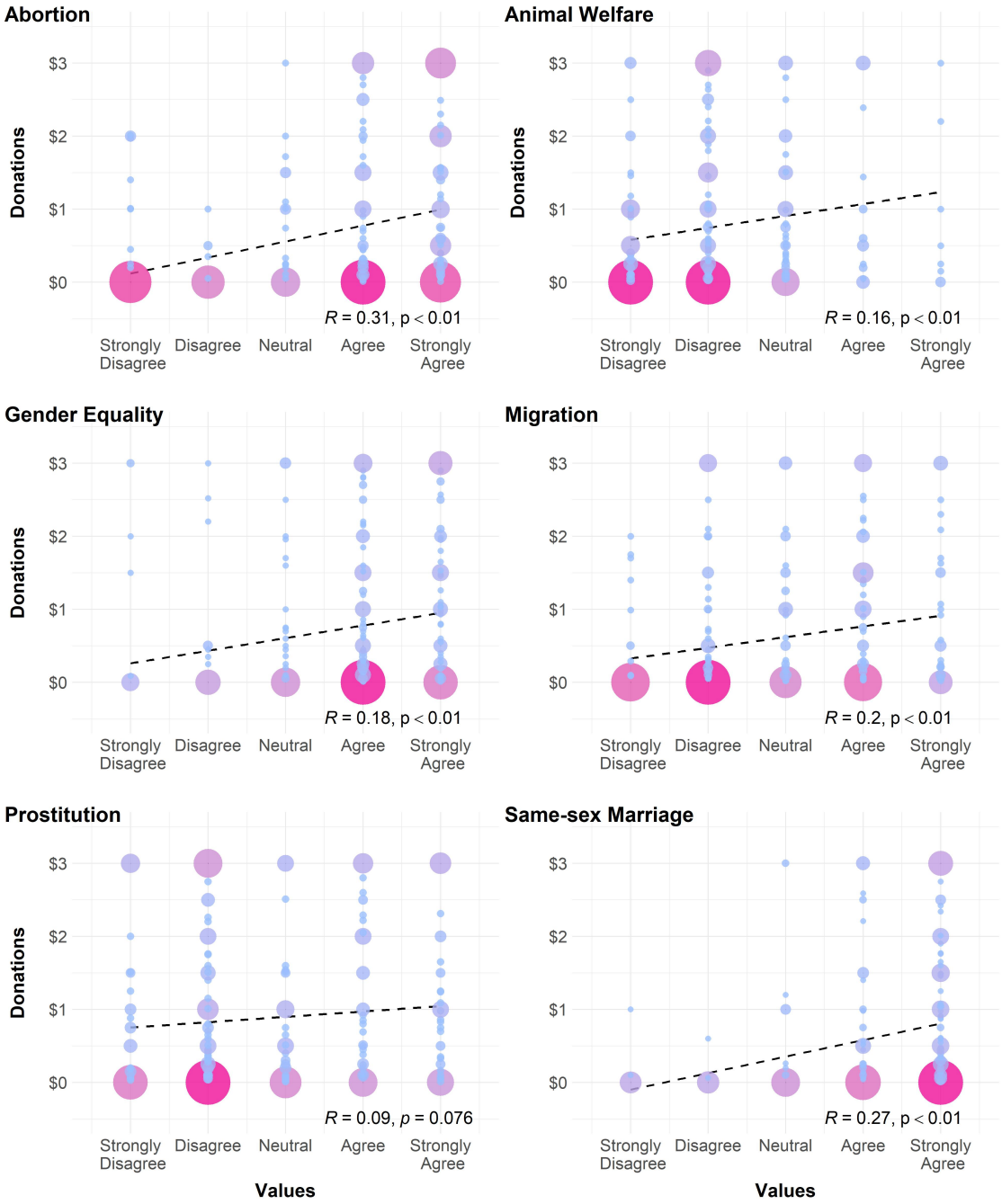
Note: The figure shows the correlation between beliefs and donations in treatment VALUENOTSALIENT separately for each of the six policy debates. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

Figure 2.A.4: Relationship between Donations and Beliefs in VALUESALIENT, by Topic



Note: The figure shows the correlation between beliefs and donations in treatment VALUESALIENT separately for each of the six policy debates. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

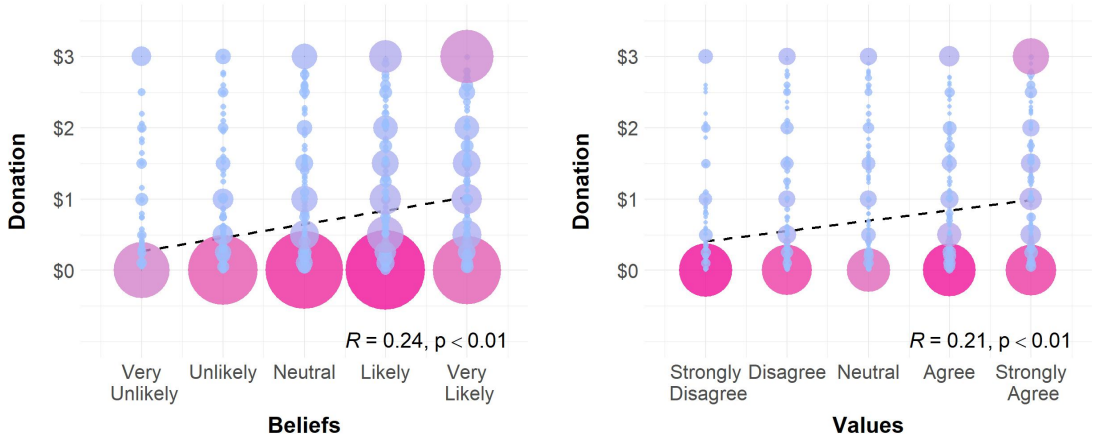
Figure 2.A.5: Relationship between Donations and Values in ValueSalient, by Topic



Note: The figure shows the correlation between values and donations in treatment ValueSalient separately for each of the six policy debates. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

2.A.2 Supplementary Results on Beliefs, Values and Donations

Figure 2.A.6: Correlation between Donations and Values/Beliefs in Treatment CONVINCE-SELF

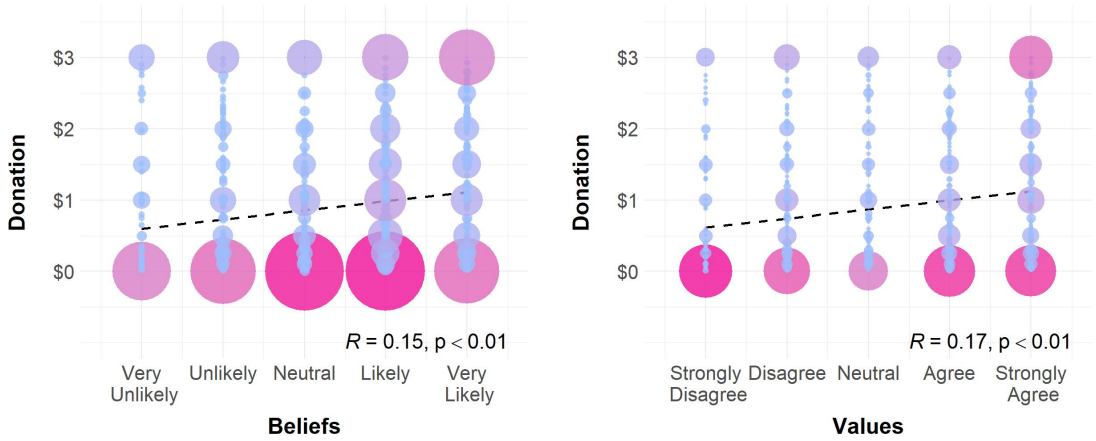


(a) Donations and Beliefs in CS

(b) Donations and Values in CS

Note: Figure 2.A.6(a) shows the correlation between beliefs and donations in treatment CONVINCE-SELF, Figure 2.A.6(b) shows the correlation between values and donations in treatment CONVINCE-SELF. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

Figure 2.A.7: Correlation between Donations and Values/Beliefs in Treatment ConvinceOther



(a) Donations and Beliefs in COR **(b)** Donations and Values in COR

Note: Figure 2.A.7(a) shows the correlation between beliefs and donations in treatment ConvinceOther, Figure 2.A.7(b) shows the correlation between values and donations in treatment ConvinceOther. The data points are weighted by the number of observations which shows in color and size of the points. The dotted line represents the result of a linear regression of donations on beliefs respectively values. The Pearson correlation coefficient, R, and its p-value are given at the bottom right of each graph.

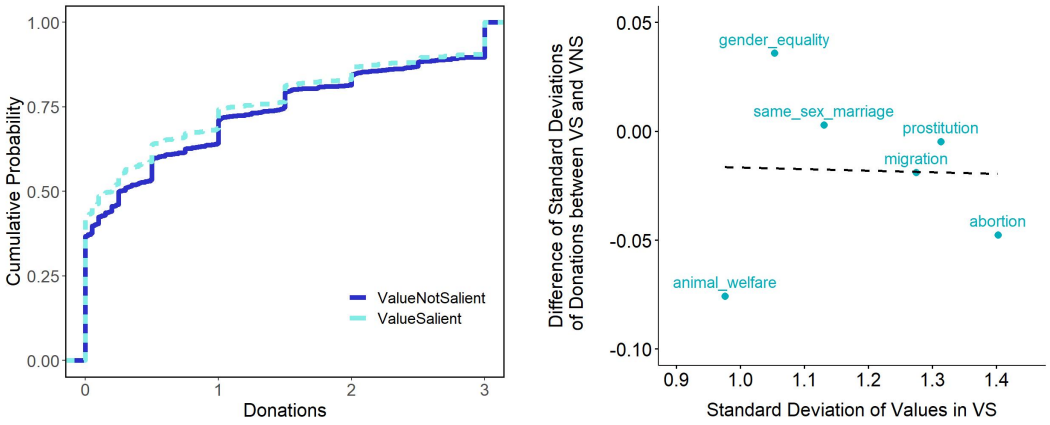
2.A.2.1 The Relationship between Hypothesis 2 and Donation Decisions

The results described in this section did not form part of our pre-registration. However, as an additional ex post analysis, we provide documentation of the relationship between the donation decisions observed in the ValueSalient and ValueNotSalient treatment conditions. The general conclusion of the results in this section is that donation decisions were not significantly impacted by the treatment variation. This indicates that although beliefs were shifted by the treatment, this shift did not translate into a change in donation behavior. This result, therefore, contributes to the growing literature that documents a complex relationship between measured beliefs and behavior. While some of the work in this literature documents evidence of beliefs causally affecting behavior in the manner predicted by standard economic models (see, e.g., Costa-Gomes, Huck, and Weizsäcker 2014; Haaland, Roth, and Wohlfart 2020; Barron and Gravert 2021), there is also body of work that show a divergence between predictions and behavior (see, e.g., Costa-Gomes and Weizsäcker 2008; Ivanov, Levin, and Niederle 2010; Haaland and Roth 2021).

In interpreting these results, it is important to keep in mind that we also observe a strong and robust correlation between donation decisions and both beliefs and values in each of our treatment conditions. There are several reasons why the shift in beliefs might not translate directly into a shift in donation decisions, including the following. First, it may be the case that deep values are a more important driver of donation decisions than factual beliefs. This would explain the correlations between donations and beliefs and values (since values and beliefs are correlated), but would also be consistent with the fact that the shift in beliefs doesn't translate into a shift in

donation decisions. Second, it is plausible that when individuals face their donation decision, the underlying contentious value debate is triggered and becomes salient at the point of making the donation decision. This would potentially negate the treatment differences generates by varying the salience of the value debates introduced by the ValueSalient and ValueNotSalient treatment conditions at the point of making the donation decision.

Figure 2.A.8: Results for Hypotheses 2a and 2b Looking at Donations

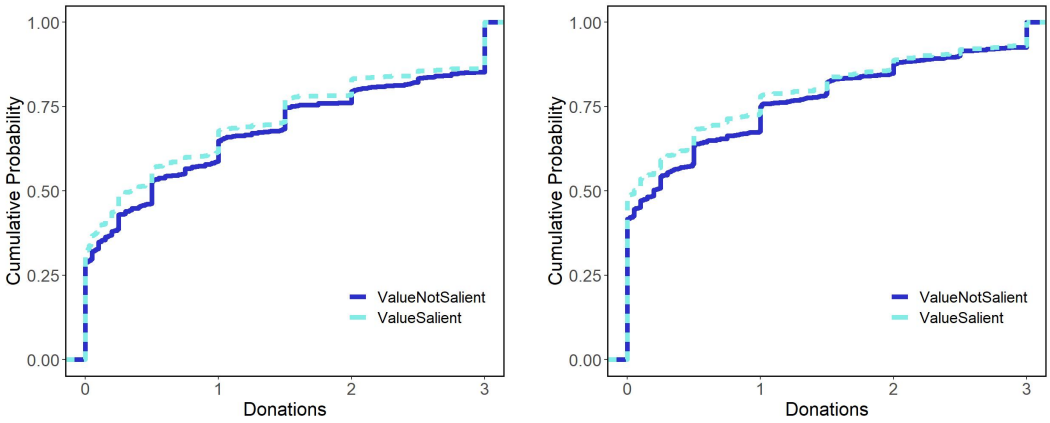


(a) CDF of Beliefs and Values in VS and VNS

(b) Change in Variance of Beliefs as a Function of Value Spread

Note: Figure 2.3(a) shows the results for Hypothesis 2a, i.e. the cumulative density function of donations in treatments ValueSalient and ValueNotSalient. Figure 2.3(b) shows the result for Hypothesis 2b. The y-axis shows the difference of the standard deviations of donations between treatments ValueSalient and ValueNotSalient and the x-axis shows the standard deviation of values in treatment ValueSalient. The dotted line depicts the result from a linear regression of the difference of the standard deviations on the standard deviation of values.

Figure 2.A.9: Results for Hypotheses 2a and 2b Looking at Donations



(a) CDF of Beliefs and Values in VS and VNS, subjects on the left of the political spectrum

(b) CDF of Beliefs and Values in VS and VNS, subjects on the right of the political spectrum

Table 2.A.1: Influence of Increased Salience of Values on Donations

	(1)	(2)	(3)
ValueSalient	-0.071	-0.019	-0.151
	[0.078]	[0.110]	[0.107]
Pol. Attitude ($\tilde{p}$)	0.284***	0.512***	0.306***
	[0.090]	[0.113]	[0.107]
ValueSalient	-0.013	-0.114	0.100
× **Pol. Attitude**	[0.127]	[0.156]	[0.150]
Constant	0.671***	0.453***	0.652***
	[0.055]	[0.080]	[0.077]
Observations	4560	2550	3006
Pol. Attitude ($\tilde{p}$) Variable	Left-Right Scale (Left = 1)	Party Affiliation (Democrat = 1)	Last Election (Clinton = 1)

Clustered standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: ValueSalient is a dummy variable equal to one if the individual was assigned to treatment ValueSalient and hence equal to zero if the individual was assigned to treatment ValueNotSalient. We use three measures of the political attitudes variable. This is indicated in the last two rows of the table. In column (1), Political Attitude is a dummy equal to one if the individual is below the median on a 1 to 10 scale of political attitudes where 1 is the most left and 10 is the most right attitude. In column (2), Political Attitude is a dummy equal to one if the individual identifies as a Democrat rather than as a Republican and in column (3) Political Attitude equals one if the individual indicated that they voted for Clinton in the 2016 elections and zero if they voted for Trump.

2.B Sample Balance

Table 2.B.1: Sample Balance.

	Full	Treatment				
	Sample	(1)	(2)	(3)	(4a)	(4b)
Age	44.09	44.10	43.82	45.04	44.49	42.97
	(15.762)	(15.874)	(15.656)	(15.706)	(15.770)	(15.820)
Female	0.52	0.51	0.51	0.52	0.51	0.54
	(0.500)	(0.501)	(0.500)	(0.500)	(0.501)	(0.499)
Etnicity						
Asian	0.06	0.07	0.06	0.07	0.06	0.06
	(0.245)	(0.254)	(0.232)	(0.249)	(0.244)	(0.244)
Black	0.13	0.16	0.12	0.11	0.14	0.12
	(0.336)	(0.365)	(0.331)	(0.308)	(0.345)	(0.330)
Mixed	0.03	0.02	0.03	0.02	0.02	0.05
	(0.171)	(0.136)	(0.181)	(0.144)	(0.156)	(0.223)
Other	0.02	0.02	0.02	0.02	0.03	0.02
	(0.143)	(0.126)	(0.134)	(0.125)	(0.172)	(0.156)
White	0.76	0.74	0.77	0.79	0.74	0.74
	(0.430)	(0.440)	(0.424)	(0.408)	(0.437)	(0.442)
Employment Status						
Starting new job	0.01	0.02	0.02	0.01	0.01	0.01
within next month	(0.108)	(0.126)	(0.134)	(0.089)	(0.091)	(0.091)
Full-Time	0.47	0.47	0.46	0.49	0.47	0.44
	(0.499)	(0.500)	(0.499)	(0.501)	(0.500)	(0.496)
Not in paid work	0.21	0.20	0.19	0.22	0.23	0.21
	(0.408)	(0.401)	(0.393)	(0.417)	(0.421)	(0.411)
Other	0.04	0.03	0.05	0.05	0.03	0.05
	(0.203)	(0.176)	(0.217)	(0.219)	(0.179)	(0.217)
Part-Time	0.18	0.19	0.20	0.16	0.17	0.19
	(0.386)	(0.394)	(0.402)	(0.366)	(0.372)	(0.393)
Unemployed	0.09	0.09	0.08	0.07	0.09	0.10
(job seeking)	(0.280)	(0.280)	(0.268)	(0.258)	(0.292)	(0.303)
Observations	1863	375	385	377	363	363

Continued on next page.

Table 2.B.1 – continued from previous page

	Full	Treatment				
	Sample	(1)	(2)	(3)	(4a)	(4b)
Do not know/	0.00	0.00	0.00	0.01	0.00	0.00
not applicable	(0.046)	(0.052)	(0.000)	(0.073)	(0.000)	(0.052)
Doctorate degree	0.03	0.03	0.03	0.03	0.04	0.04
	(0.181)	(0.161)	(0.174)	(0.169)	(0.206)	(0.193)
Graduate degree	0.16	0.18	0.14	0.17	0.17	0.13
	(0.366)	(0.386)	(0.350)	(0.378)	(0.372)	(0.339)
High school diploma	0.22	0.23	0.18	0.22	0.21	0.26
	(0.414)	(0.423)	(0.386)	(0.413)	(0.407)	(0.437)
No formal	0.00	0.01	0.00	0.00	0.00	0.00
qualifications	(0.052)	(0.073)	(0.051)	(0.052)	(0.000)	(0.052)
Secondary education	0.02	0.01	0.01	0.02	0.03	0.02
	(0.130)	(0.073)	(0.113)	(0.153)	(0.164)	(0.128)
Technical/	0.17	0.16	0.19	0.15	0.17	0.18
community college	(0.376)	(0.367)	(0.395)	(0.361)	(0.374)	(0.382)
Undergraduate degree	0.40	0.39	0.44	0.40	0.39	0.37
	(0.489)	(0.488)	(0.497)	(0.490)	(0.487)	(0.485)
Income						
Less than $10000	0.06	0.07	0.05	0.07	0.07	0.06
	(0.242)	(0.250)	(0.222)	(0.249)	(0.249)	(0.239)
$10000-$15999	0.06	0.08	0.05	0.06	0.07	0.04
	(0.236)	(0.267)	(0.211)	(0.244)	(0.254)	(0.193)
$16000-$19999	0.03	0.03	0.04	0.02	0.03	0.03
	(0.171)	(0.176)	(0.200)	(0.135)	(0.164)	(0.172)
$20000-$29999	0.11	0.12	0.09	0.11	0.10	0.13
	(0.313)	(0.322)	(0.292)	(0.315)	(0.299)	(0.336)
$30000-$39999	0.10	0.11	0.12	0.11	0.09	0.10
	(0.304)	(0.309)	(0.325)	(0.308)	(0.280)	(0.296)
$40000-$49999	0.10	0.11	0.08	0.09	0.12	0.11
	(0.303)	(0.309)	(0.268)	(0.291)	(0.330)	(0.314)
$50000-$59999	0.10	0.09	0.10	0.10	0.10	0.10
Observations	1863	375	385	377	363	363

Continued on next page.

Table 2.B.1 – continued from previous page

	Full	Treatment				
	Sample	(1)	(2)	(3)	(4a)	(4b)
	(0.298)	(0.288)	(0.302)	(0.294)	(0.303)	(0.303)
$60000-$69999	0.07	0.06	0.08	0.09	0.07	0.06
	(0.254)	(0.230)	(0.268)	(0.283)	(0.249)	(0.234)
$70000-$79999	0.08	0.09	0.09	0.08	0.09	0.07
	(0.273)	(0.280)	(0.280)	(0.267)	(0.284)	(0.254)
$80000-$89999	0.05	0.05	0.05	0.04	0.05	0.05
	(0.214)	(0.208)	(0.222)	(0.202)	(0.223)	(0.217)
$90000-$99999	0.05	0.04	0.05	0.04	0.05	0.05
	(0.208)	(0.196)	(0.211)	(0.196)	(0.223)	(0.212)
$100000-$149999	0.11	0.10	0.13	0.12	0.09	0.13
	(0.319)	(0.306)	(0.334)	(0.331)	(0.280)	(0.339)
More than $150000	0.05	0.04	0.05	0.05	0.06	0.04
	(0.218)	(0.196)	(0.227)	(0.224)	(0.244)	(0.193)
Prefer not to say	0.03	0.03	0.02	0.02	0.02	0.04
	(0.162)	(0.176)	(0.151)	(0.144)	(0.138)	(0.193)
Prolific Score	99.58	99.59	99.67	99.60	99.52	99.52
	(1.366)	(1.242)	(1.158)	(1.276)	(1.682)	(1.430)
Left/Right Attitude	4.60	4.76	4.71	4.52	4.42	4.58
	(2.552)	(2.556)	(2.449)	(2.621)	(2.537)	(2.594)
Observations	1863	375	385	377	363	363

Note: The table shows the means and standard deviations (in parenthesis) of demographic variables for the individuals in our sample. The first column provides the information on the whole sample, column 2 (Treatment (1)) looks at treatment ValueNotSalient, column 3 (Treatment (2)) at treatment ValueSalient, column 4 (Treatment (3)) at treatment ConvinceSelf, column 5 (Treatment (4a)) at treatment ConvinceOther, column 6 (Treatment (4b)) at treatment BeingConvinced.

2.C Preregistration Document

Morals, Beliefs, and Actions

Kai Barron (WZB), Anna Becker (UCL), Steffen Huck (UCL and WZB)

This version: 14/01/2020

PART I: EXPERIMENTAL DESIGN

Setup of Experiment

This experiment will be run as an online survey. The sample for the four main treatments will consist of 1,500 individuals that are representative of the US population in terms of age, sex, and ethnicity. An additional 375 subjects will be recruited later for an auxiliary treatment, Treatment 4b, which builds on subjects' choices in Treatment 4a. All participants will be recruited via the online platform Prolific. Subjects will be paid £3 for participation and have the option to earn a bonus in Treatments 1, 2, 3 and 4b.[20] A strict no-deception policy will be followed. The experiment is programmed using the experimental software o-Tree (Chen, Schonger, and Wickens (2016)).

Experimental Design

At the beginning of the experiment, subjects are randomized into one out of four treatment groups such that each group consists of 375 subjects. The four treatments are described in the following text.

20. The show-up fee is converted into US dollars, since the subjects are recruited from the USA.

Treatment 1 – "Control"

Part 1

Subjects are asked to state how likely they think a statement that they are presented with is true. The statements have been chosen such that they can be associated with a policy domain and typically refer to facts on which scientific consensus has not been reached yet. Subjects use a five-point Likert scale to indicate their beliefs. They can choose between the following options: "Very Unlikely", "Very Likely", "Neutral", "Likely" and "Very Likely". After a waiting time of 15 seconds subjects can proceed to the next page.

This is repeated six times for six different policy domains. These are migration, animal welfare, gender equality, abortion, prostitution and gay rights. Table 2.C.1 in the Appendix provides an overview over all domains and the statements presented to subjects. The order with which the statements on the different domains are shown to participants is randomised at the individual level.

Part 2

After subjects have submitted their beliefs on the six different issues, they are informed that they have the option to make six donations to charities. They are also informed that one out of the six decisions they are about to make will be chosen at random to be implemented.

Subjects then see the statement they were earlier confronted with, the belief that they stated and the option to donate to a charity that is active in the respective policy domain. As in Part 1, this process is repeated six times following the same order of

domains as in Part 1. Subjects are informed briefly about the aims of the charities and can indicate with a slider how much they would like to donate. They are provided with $3 and can choose to donate any amount between $0 and $3. Subjects will be paid the remaining amount (i.e. what they decided not to donate) at the end of the experiment. They can proceed to the next page at any time after a waiting time of 15 seconds.

Treatment 2 – "Moral Values"

Part 1

Treatment 2 is similar to Treatment 1, with the following exceptions.

Different to Treatment 1, in Treatment 2 subjects will also be asked to state some moral values that are related to the same six policy domains. The question of how much they agree or disagree with a moral statement they are presented with appears above the question regarding the factual statement. Subjects use a five-point Likert scale to indicate their agreement. They can choose between the following options: "Strongly Disagree", "Disagree", "Neutral", "Agree" and "Strongly Agree". The moral statements can be found in Table 2.C.1 in the Appendix.

Part 2

Different to Treatment 1, in Treatment 2, subjects are also reminded of the moral values they stated in Part 1. The screen will therefore show the moral statement and the subject's choice above the factual statement and the subject's decision and then offers the subjects to donate to a proposed charity.

Treatment 3 – "Convincing Yourself"

As in Treatment 2, subjects are asked to state moral values, factual beliefs, and then make a charitable donation. The key difference between Treatment 3 and Treatment 2 is that in Treatment 3 the moral statement and the factual statement are presented to subjects on the same screen as the option to donate to charity, which is presented at the bottom of the page. They receive the same information as subjects in Part 2 of Treatments 1 and 2 and face the same charitable giving decision, but in Treatment 3 all three decisions are made on the same page (i.e. the beliefs, moral value judgments, and charitable donations).

As in the other treatments, this will be repeated six times for the six different policy domains where the order is randomized on the individual level. Subjects use a Likert scale to indicate their values and beliefs and a slider to indicate how much they would like to donate. They are provided with $3 per decision. One of those decisions will be chosen at random to be implemented of which subjects are informed about in advance. They will be paid what they decide not to donate after the experiment.

Treatment 4a – "Convincing Others"

As in Treatment 2, subjects are asked to state their moral values and their factual beliefs. The moral statement and the factual statement are presented to subjects on the same screen underneath each other as in Part 1 of Treatment 2.

Before stating values and beliefs, the subject is informed that another participant will have the option to donate to a related charity. Importantly, the other participant will make their charitable donation decision after being informed about the moral values

and factual beliefs that the subject reports (i.e. the moral values and factual beliefs reported by subjects in Treatment 4a will be sent to subjects in Treatment 4b before participants in 4b make their charitable donation decisions).

Subjects in Treatment 4a are presented with the information the other participant in Treatment 4b will be shown about the charity. The donation decision that will be completed by Treatment 4b subjects is the same as in the previous Treatments, i.e. the participant has $3 available of which they keep what they decide not to donate. Both the subjects in Treatment 4a and the other participant in Treatment 4b will be informed in advance that one of the six decisions will be chosen at random to be implemented.

After the main Treatments 1, 2, 3 and 4a have been run, another 375 subjects will be recruited for Treatment 4b:

Treatment 4b – "Being Convinced"

Part 1

This will be identical to Part 1 of Treatment 2.

Part 2

In Treatment 4b, Part 2 will be similar to Treatment 2, with the exception that instead of subjects being reminded of their own decisions regarding the moral and the factual statements, subjects are now informed about the decisions of a participant from Treatment 4a when they make their charitable donation decision. As before, subjects will be provided with $3 for each decision of which they will be paid what

they decide not to donate. They are also informed that one out of their six decisions will be chosen at random to be implemented.

Post-experimental Survey

After the experiment, all subjects will be asked to fill out a survey. The survey covers the following topics:

1. Personal Details
2. Political Attitude
3. Religious Attitude
4. Moral Foundations Questionnaire[21]
5. Questions on Moral Behavior
6. Cognitive Reflection Test

PART II: ANALYSIS PLAN

II.1) Introduction

Standard theories on belief formation typically disregard the desire of individuals to gather, avoid or interpret information in a way that serves non-instrumental purposes. A large recent literature has, however, shown that, for example, self-serving biases, wishful thinking and motivated reasoning are important determinants of belief formation. This project studies individuals' moral values as a potential source for

21. https://moralfoundations.org/questionnaires/, accessed 31/12/2019.

motivated cognition and links it to partisan disagreement about factual statements, i.e. polarisation.[22]

We proceed in three steps. In each step, we test a small set of hypotheses that are interlinked by a common underlying idea. In the first step, we seek to test whether individuals report moral values and factual beliefs that are aligned in the domain of political issues. Potential reasons why individuals might do this include: i) avoiding emotional discomfort, or cognitive dissonance emanating from holding or stating incoherent values and beliefs, and ii) using value and belief statements to justify self-interested actions (e.g. actions that increase the individual's material wealth). From the analysts' perspective, the presence of such belief-value constellations would provide a basis for taking an individual's moral values into consideration in trying to understand belief formation regarding factual statements. A potential motive for this could be a desire to establish something akin to a "moral identity" (Bénabou and Tirole (2011)).

In the second step, we then test whether there is a systematic pattern in the way factual beliefs are formed that may be partially responsible for the creation of these belief-value constellations (i.e. we ask whether factual beliefs are constructed in a way that forms these belief-value constellations). To do this, we compare the distribution of beliefs of individuals that were previously asked about, and hence reminded of, their related values (Treatment 2) with the distribution of beliefs of individuals in the control Treatment 1 (where no moral value statements are reported prior to stating factual beliefs). This allows us to assess the influence of being primed to think about the particular factual belief in question through the lens of the related value

22. See for example Rabin (1995) for a theory on self-serving biases in moral reasoning.

debate. Following on from this, we ask whether this mechanism can explain the recent trend of a polarization of beliefs in society that has been demonstrated to run along ideological lines (see, e.g., Gentzkow (2016)). In particular, there appears to be an increased disagreement about objective facts among members of society that is associated with political attitudes.[23] We hypothesize that the heterogeneity in moral values between different political groups may be leading to the formation of these polarized factual beliefs. Hence, we test whether factual beliefs become polarized as a function of political attitudes (e.g. as a result of individuals' desire to adjust their beliefs to their moral values.)[24]

Lastly, we study two potential forces that might increase or decrease the degree of polarization of (stated) beliefs. First, we look at the impact of financial incentives (through motivated reasoning or self-persuasion), and second we will study the role of persuading others, which is particularly relevant in political contexts.

The first channel which considers the role of financial incentives on belief formation is important because there are many reasons why individuals might face costs to hold certain beliefs or values. For example, it may be costly to hold different beliefs and values to those held by individuals in one's peer network.[25] Alternatively, holding particular beliefs and values may be costly when they induce the individual to take a particular costly action. For example, an individual who advocates the merits

23. The most prominent current example is probably that of climate change where there is a widening gap in the views on the scientific evidence between Republicans and Democrats in the US (see e.g. McCright and Dunlap (2011)).

24. In his theoretical work, Le Yaouanq (2021) links heterogeneity in political attitudes to partisan disagreement about objective facts through people's idiosyncratic preferences regarding the policy implications of scientific findings. Our work seeks to understand the underlying psychological mechanisms in more detail.

25. On the role of group identity in belief polarization see e.g. Gennaioli and Tabellini (2018).

reducing inequality in society may feel compelled to take actions that reduce their own wealth in order to increase the wealth of a poorer individual. Rather than studying abstract costs (e.g. incoherence with one's peers' beliefs), we focus on the latter type of costs that accrue due to taking actions that reduce one's personal payment. In particular, we consider costly donation decisions and hypothesize that this cost leads individuals to bias their (stated) beliefs and values when they expect a related donation decision, with the bias operating in the opposite direction to the beliefs and values consistent with a higher donation.[26]

Second, we look at the potential role of individuals' desire to convince others to take actions that are in line with their own values. This motive of convincing others might lead people to further align their (stated) beliefs with their political agenda or goals, and perhaps to exaggerate these stated beliefs. We study whether subjects adjust their beliefs to be more extreme in order to convince another participant to give more or less to a suggested charity. In this case, we test whether individuals overstate the strength of their beliefs when trying to persuade another person to act in a certain way.

Section II.2) introduces the necessary notation, before we formalize our hypotheses in Section II.3).

2.C.1 Notation

Let b_t denote the factual beliefs stated by individuals in Treatment t, where $t \in \{1,2,3,4a,4b\}$, v_t are the moral values stated by individuals in treatment t where

26. It is also possible that there are individuals that now bias their values and beliefs in the same direction as beliefs and values consistent with a higher donation as they consider them to be more important when relevant to justify a charitable donation. Our prior, however, is that the effect described above dominates.

$t \in \{2,3,4a,4b\}$ and d_t are the donation decisions of individuals in treatment t where $t \in \{1,2,3,4b\}$. Let F_{b_t} denote the cumulative distribution function (cdf) of factual beliefs in Treatment t where $t \in \{1,2,3,4a,4b\}$, F_{v_t} the cdf of moral values in Treatment t where $t \in \{2,3,4a,4b\}$, and F_{d_t} the cdf of donations in Treatment t where $t \in \{1,2,3,4b\}$.

Let p_t denote the left-right political stance of individuals in Treatment t which will be elicited for all participants in the post-experimental survey using a Likert scale ranging from 1 to 10 where 1 is left and 10 is right, i.e. p_t is increasing in the degree to which an individual positions herself on the right of the political spectrum. F_{p_t} denotes the respective cdf. Belief and value statements were chosen such that the factual statement being true would provide support for agreement to the value statement. All statements are coded this way for the analysis but not necessarily presented to participants like this (Table 2.C.1 in the Appendix shows how statements are presented to subjects during the experiment).

At the same time, we recode all the moral value variables such that they are likely to be increasingly appealing as one moves from the right to the left of the political spectrum. Similarly, we code the factual belief variables such that if they are true, they support moral value positions typically held by individuals on the political left.[27] Charities are chosen such that if the moral statement is supported it would justify the charities' objectives.[28]

27. The policy domain "Prostitution" is an ambiguous case. Individuals from the political left could both support and reject more liberal prostitution rights.

28. Therefore, it is worth pointing out that variables are coded such that a higher value of b_t, v_t, and d_t should be consistent with a lower value of p_t according to the researcher team's priors.

2.C.2 Hypotheses

2.C.2.1 Belief-Value Constellations

As stated above, we begin by testing whether individuals report moral values and factual beliefs on political issues that are aligned, whether their moral values are aligned with their political attitudes and whether factual beliefs and moral values are related to decision making in the form of costly charitable donation choices.

Hypothesis 1

a) Moral values are positively correlated with beliefs:

$$Corr(v_2, b_2) \geq 0.$$

b) Moral values are negatively correlated with political attitudes:

$$Corr(v_2, p_2) \leq 0.$$

c) Donations are positively correlated with beliefs and values:

$$Corr(d_1, b_1) \geq 0, Corr(d_2, b_2) \geq 0, Corr(d_2, v_2) \geq 0.$$

Part a) of Hypothesis 1 tests whether moral values and factual beliefs reported by subjects in Treatment 2 are aligned. Recall that in Treatment 2 subjects are not aware of the opportunity to donate by the time they state their values and beliefs.

Part b) tests for a negative correlation between moral values and political attitudes.

For example, Enke, Rodriguez-Padilla, and Zimmermann (2016) show that an individuals' moral type is strongly correlated with their political affiliation. Rather than looking at predefined moral types we look at concrete moral convictions with regard to certain policy domains. We expect that the further an individual is on the left of the political spectrum (i.e. the lower is p_2) the more likely they are to agree with the moral value statement.

In Part c) of Hypothesis 1, we hypothesize that individuals donate more when their moral values and their beliefs are such that the cause of the charity is justified by them (i.e. that donations are positively correlated with beliefs and moral values consistent with the charity's mandate).

2.C.2.2 Construction of Beliefs

The following hypothesis tests: (i) whether the formation of factual beliefs is influenced by the values individuals hold, thereby explaining the formation of belief-value constellations, and (ii) whether this influence of values on belief formation may lead to a polarization of factual beliefs across the political spectrum.

Hypothesis 2

a) The distribution of factual beliefs in Treatment 1 is different from the distribution of factual beliefs in Treatment 2:

$$F_{b_1} \neq F_{b_2}.$$

b) Comparing across the six domains indexed by m, the difference between the variance in beliefs in Treatment 2 and the variance in beliefs in Treatment 1 is

increasing in the variance in values in Treatment 2:

$$\frac{d[Var(b_2^m) - Var(b_1^m)]}{d[Var(v_2^m)]} \geq 0.$$

c) Beliefs in Treatment 2 are more polarized than beliefs in Treatment 1:

$$E(b_2|p_2 < E(p_2)) - E(b_1|p_1 < E(p_1)) \geq E(b_2|p_2 > E(p_2)) - E(b_1|p_1 > E(p_1)).$$

Part a) of Hypothesis 2 exploits the fact that unlike subjects in Treatment 2, subjects in Treatment 1 are not asked to state their moral values. We test whether reminding subjects of their moral values has an impact on the distribution of factual beliefs which would indicate that moral values are relevant for the construction of beliefs. In Part b) of Hypothesis 2, we go further and test whether values exert a systematic influence on belief formation. In particular, Part b) of Hypothesis 2 posits that when there is more dispersion in the values that subjects hold with regard to a certain policy domain, we expect that there will be a shift towards more extreme beliefs as subjects are drawn towards more coherent belief-value constructions.

The last part of Hypothesis 2 tests whether a polarization in beliefs can be explained as a result of the posited impact of values on belief formation. The inequality in Part c) states that the difference in the means of beliefs between Treatment 2 and Treatment 1 is greater for subjects below the mean of political attitudes, i.e. relatively on the left of the political spectrum, than for subjects above the mean of political attitudes, i.e. relatively on the right of the political spectrum. To put this another way, the hypothesis states that individuals on the left will increase their beliefs between

Treatment 1 and Treatment 2 more than individuals on the right of the political spectrum, on average. Note that this also includes the case where subjects adjust their beliefs downwards (i.e. it is completely consistent with individuals on the right shifting their beliefs downward between Treatment 1 and 2, which would make the right-hand side negative).

If political attitudes are sufficiently widely dispersed, we would expect a positive left-hand side and a negative right-hand side, i.e. what we traditionally refer to as polarization. Otherwise, we expect to see a mild form of polarization where beliefs are adjusted to different extents by those on the left and the right of the political spectrum within our sample.

2.C.2.3 Convincing Yourself and Convincing Others

The last hypothesis is split in two parts. In Part A, we look the role of self-serving biases that allow subjects to justify selfish behaviour and are expected to lead to a downward[29] bias in beliefs, values and charitable donations. Part B, on the other hand, studies whether introducing the opportunity to convince another participant to take an action that is in line with one's moral values can lead to a greater polarization of beliefs.

Hypothesis 3

A. **Convincing Yourself**

a) When there is an increase in the cost of holding certain beliefs and values individuals may shift their stated beliefs and values in the opposite

29. Here, "downward" refers to a bias in the direction that is consistent with being self-serving. In terms of the way we have defined our variables, it will also refer to lower values of b_t, v_t, and d_t.

direction:

i) b_2 first-order stochastically dominates b_3, i.e. $F_{b_2} \leq F_{b_3}$.

ii) v_2 first-order stochastically dominates v_3, i.e. $F_{v_2} \leq F_{v_3}$.

b) Donations in Treatment 3 are lower than in Treatment 2:

$$E(d_2) \geq E(d_3).$$

B. **Convincing Others**

c) Beliefs in Treatment 4a are polarized in comparison to beliefs in Treatment 2:

$$E(b_{4a}|p_{4a} < E(p_{4a})) - E(b_2|p_2 < E(p_2)) \geq$$
$$E(b_{4a}|p_{4a} > E(p_{4a})) - E(b_2|p_2 > E(p_2)).$$

In Treatments 1, 2 and 3 subjects are given the opportunity to donate to charity. Donating comes at the cost of a foregone bonus payment. Previous work has shown that people develop self-serving biases in order to excuse their selfishness in charitable giving (see e.g. Exley (2015) on the role of risk or Exley (2020) on using charity performance metrics as an excuse). In Part a) of Hypothesis 3 we test whether subjects shift their values and beliefs downwards to justify smaller donations in order to receive higher bonuses.

In Part b), we hypothesize that individuals donate less on average in Treatment 3

than in Treatment 2. In both treatments, a higher donation reduces the payment that the subject receives herself (i.e. the material incentives are identical across the two treatments). However, in Treatment 2, individuals only learn about the possibility to donate after they have already stated their values and beliefs which rules out the opportunity to make any adjustments to one's values in order to justify self-interested charitable donation decisions.[30]

The last part of Hypothesis 3 (i.e. Part c) tests whether individuals report more polarized beliefs when they have the opportunity to convince someone to make a donation which is the case in Treatment 4a. As in Hypothesis 2 c), the inequality in Hypothesis 3c) states that the difference in the means of beliefs between Treatment 4a and Treatment 2 is greater for subjects below the mean of political attitudes, i.e. relatively on the left of the political spectrum, than for subjects above the mean of the political spectrum, i.e. relatively on the right of the political spectrum.

As a spillover from Hypothesis 3 and Treatment 4a we can also study the effect of persuasion on recipients of the persuasion messages by looking at Treatment 4b. More specifically, we can study whether donations to charity increase when an individual sees their own values and beliefs confirmed by another person. We would expect that there will be a polarization of donation decisions when the political attitudes of the sender and the receiver are aligned.[31] In cases where the sender's and the

30. There is the possibility that there exists a subset of individuals in Treatment 3, who instead of shifting down their reported beliefs and values to justify a lower charitable donation decision, instead shift up their values and beliefs in order to then enhance the signalling value of their donations. These individuals might then also donate more when facing Treatment 3 in comparison to Treatment 2. This effect would operate in the opposite direction to the main effect hypothesised in Hypothesis 3.b). For simplicity, we have stated Hypotheses 3.b) in terms of the average effect for the entire sample, working under the assumption that the main hypothesised effect of adjusting one's beliefs and values downwards in a self-serving fashion will dominate this potential countervailing secondary effect.

31. This could be tested similarly to Part c) of Hypotheses 2 and 3.

receiver's political attitude are not aligned, we can think of (at least) two opposing possible effects. Individuals might either doubt their own convictions and reassess them or they might want to prove the sender wrong by exaggerating or lowering the donated amount. Ex ante, it is unclear which effect is expected to dominate. This in turn, might also depend on political affiliation. We do not propose any hypothesis here, but will document the data in an exploratory analysis.

2.C.3 References for Preregistration Document

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

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree – An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9:88–97.

Enke, Benjamin, Ricardo Rodriguez-Padilla, and Florian Zimmermann. 2016. "Moral Universalism and the Structure of Ideology." *Working Paper*.

Exley, Christine L. 2015. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies* 83, no. 2 (October): 587–628.

———. 2020. "Using Charity Performance Metrics as an Excuse Not to Give." *Management Science* 66 (2): 553–563.

Gennaioli, Nicola, and Guido Tabellini. 2018. *Identity, Beliefs, and Political Conflict.* Technical report. Working Paper.

Gentzkow, Matthew. 2016. *Polarization in 2016.* Technical report. Toulouse Network for Information Technology Working Paper.

Le Yaouanq, Yves. 2021. *Motivated cognition in a model of voting.* Technical report. Working Paper.

McCright, Aaron M., and Riley E. Dunlap. 2011. "The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010." *Sociological Quarterly* 52 (2): 155–194.

Rabin, Matthews. 1995. *Moral Preferences, Moral Constraints, and Self-Serving Biases.* Technical report. Berkeley Department of Economics Working Paper No. 95-241.

2.C.4 Appendix to Preregistration Document

Table 2.C.1: Overview over Statements and Charities.

	Debate	**Moral Statement**	**Factual Statement**	**Donation**
		"How much do you agree with the following statement?"	*"How likely do you think it is that the following statement is true?"*	Charity
1	Migration	People should be allowed to migrate freely between countries.	All countries benefit economically from the free movement of labour.	American Immigration Council
2	Animal Welfare	It is wrong to eat animals.	Animals feel less pain than humans.	World Animal Protection
3	Gender Equality	Gender equality should be an objective of policy-making.	Discrimination against women is the primary reason why women earn less than men.	Equality Now
4	Abortion	Abortion should be legal.	Women who have had an abortion experience more psychological distress than women who have had a miscarriage.	Planned Parenthood
5	Prostitution	Prostitution should be illegal.	Human trafficking is facilitated by liberal prostitution laws.	A21
6	Same-sex Marriage	Gay couples should have the same rights as heterosexual couples.	Societies where same-sex marriage is legal are happier than societies where it is illegal.	OutRight

Table 2.C.2: Description of Charities.

	Debate	**Charity**	**Text to introduce charity in experiment**
1	Migration	American Immigration Council	The American Immigration Council envisions an America that values fairness and justice for immigrants and believes that immigrants are part of the national fabric, bringing energy and skills that benefit all Americans. To advance change they engage in litigation, research, legislative and administrative advocacy, and communications.
2	Animal Welfare	World Animal Protection	World Animal Protection works towards a world where animals live free from suffering. They seek to improve the living conditions of animals farmed for food, to protect and save wild animals, animals affected by disasters, and working animals.
3	Gender Equality	Equality Now	Equality Now believes in creating a just world where women and girls have the same rights as men and boys. They use a unique combination of legal advocacy, regional partnership-building and community mobilization to encourage governments to adopt, improve and enforce laws that protect and promote the rights of women and girls.
4	Abortion	Planned Parenthood	The mission of Planned Parenthood is to provide comprehensive reproductive and complementary health care services in settings which preserve and protect the essential privacy and rights of each individual and to advocate public policies which ensure access to such services. They provide information and support to women considering to end a pregnancy in a clinic or using an abortion pill.
5	Prostitution	A21	The mission of A21 is to end human trafficking and slavery. They work closely with law enforcement on the ground to support police operations, identify victims through their hotlines, assist in the prosecution of traffickers, and represent survivors in court proceedings.
6	Same-sex marriage	OutRight	OutRight envisions a world where LGBTIQ (lesbian, gay, bisexual, transgender/transsexual, intersexual and queer) people everywhere enjoy full human rights and fundamental freedoms. They seek to fill research gaps, provide trainings to community members and allies to develop their expertise, and convene key stakeholders to exchange information on best practises related to ending violence based on sexual orientation.

Chapter 3

The Social Foundations of Xenophobia *(with Hans-Joachim Voth)*

Abstract

We examine the link between skewed sex ratios and xenophobic attitudes. To do so, we look at Germany where some regions have a striking surplus of men, especially among the younger generations. At the same time, the country experienced a substantial increase in anti-refugee attacks following the arrival of more than one million refugees between 2015 and 2017. In a first step, we show that higher male-to-female ratios are associated with more anti-immigrant violence and far-right voting in Germany. We then corroborate these correlations using an instrumental variables strategy exploiting the close-by availability of university places for females.

3.1 Introduction

Since the Fall of the Berlin Wall, more than 1.5 million primarily young adults have left East Germany towards the West. Since then, the number of those leaving the East has been exceeding the one of those moving there each year. These total numbers, however, disguise an important phenomenon.[1] Since 1991, the number of women leaving East Germany is far higher than that of men moving to the West. This has left East Germany with a shortage of women which is unparalleled in Europe. The economic impact of unbalanced gender ratios has been analyzed previously (see e.g. Teso 2018; Goldin and Olivetti 2013; Heer and Grossbard-Shechtman 1981, for the effects on female labour force participation). Little is known, however, about the consequences of this kind of gender-specific migration on those who are left behind empty-handed. Young men in East Germany are not only struggling in education and in the labour market.[2] They are also substantially more likely to be single and to live alone than their West German counterparts.[3][4] It is unlikely that it leaves these men unaffected when their prospect of finding a partner is so much lower than for previous generations and other men of their age. Basic biological and emotional needs cannot be fulfilled and expectations on life are not met.

In this paper, we demonstrate that skewed sex ratios can lead to increased outgroup discrimination. We hypothesize that a high level of male-male competition for female

1. The phenomenon was first discussed in detail by Kröhnert and Klingholz (2006).

2. This is not only true for Germany but also in international comparison. We see that women outperform men in education (Becker, Hubbard, and Murphy 2010) and have improved their relative standing in the labour market over time (Olivetti and Petrongolo 2016).

3. In 2000, 37 percent of men between the age of 25 and 34 are without a partner, whereas it is 46 percent for the same age group in East Germany (Kröhnert and Klingholz 2006).

4. More recent data for the US suggests, that sexlessness amongst young men has increased dramatically. Data from the GSS shows that in 2008 only 10 percent of men under the age of 30 reported having had no female sex partners since they turned 18; this share rose to 28 percent in 2018 (Ingraham 2019).

partners increases the tendency of men to perceive other - especially foreign - males as a threat in the competition for women. We examine the Eastern part of Germany which has one of the highest rates of gender imbalance on earth, with males outnumbering females in some age groups by 15:10; it is also a hotbed of support for far-right parties, with frequent attacks on foreigners. We first show that there exists a strong association between male-to-female ratios and anti-immigrant crimes and vote shares for an extreme right-wing party. We then use an instrumental variables strategy to show that the effect is arguably causal.

In a first step, we use information on gender ratios in the 401 German districts, our level of analysis, together with other demographic characteristics of the districts to establish a correlation between gender ratios and occurrences of hate crime and second votes obtained by the AfD in the 2017 elections.[5] A higher share of women is associated with less hate crime and fewer votes for the AfD. As the results from our simple regression approach are likely to be biased we confirm them with an instrumental variable approach. We use information on the share of female students enrolled in different subject groups together with data on student numbers at all German universities to compute a measure for the share of student places typically taken up by women in each district. Our instrument is strongly correlated with the gender ratio and our IV results confirm the previously documented relationship. One more woman per 100 men lowers the vote share for the AfD by about 0.8 percentage points and the incidence of hate crime by about 0.0007 percentage points (about 18 percent of the mean). Our results therefore constitute a first pass for our hypothesis

5. Every voter in Germany has two votes. With the first vote they vote for a candidate from their electoral district where the winner is determined in a first-part-the-post system. The second vote is cast for the list of a party. The overall share of second votes determines the number of seats they obtain in the German parliament.

that male-male competition could be a driver of xenophobic sentiments.

The literature on the drivers of anti-immigrant sentiments has been growing over the last years as these are believed to explain a large part of the electoral success of populist far-right parties and politicians. Several studies look at the effect that (refugee) migration has on voting outcomes (see e.g. Halla, Wagner, and Zweimüller 2017; Barone et al. 2016; Dustmann, Vasiljeva, and Piil Damm 2018; Medez and Cutillas 2014; Otto and Steinhardt 2014; Edo et al. 2019).[6] They find that a higher share of refugees is associated with an increase in the vote share for center-right and far-right parties but fail to explain the source of the xenophobic attitudes that drive these results. Other studies focus more on the economic and cultural concerns which can explain negative attitudes towards migrants. Falk, Kuhn, and Zweimüller (2011) show that right-wing criminal activities are positively related to unemployment. Dippel et al. (2021) find that the support for right-wing parties in Germany rises with the exposure to imports from low-wage countries. Cantoni, Hagemeister, and Westcott (2019) show that votes for the AfD in Germany can partly be explained by long-run cultural persistence of right-wing ideology.[7] Using survey results, Bansak, Hainmueller, and Hangartner (2016)[8] find that preferences over asylum seekers depend on their potential contribution to the economy, their deservingness and are also driven by a strong bias against Muslims. This study provides a novel perspective on the subject as it explores the underlying biological and social foundations of xenophobia. Unlike previous work, we seek to understand the social mechanisms which

6. Steinmayr (2021) studies Upper Austria and finds that short-term exposure to refugees can increase far-right voting, whereas sustained contact might actually decrease it.

7. They find that the persistence of far-right voting is due to the persistence of certain conservative values and not linked to Nazi support or persistent patterns of antisemitism as initially documented by Voigtländer and Voth (2012) and Voigtländer and Voth (2015).

8. Müller and Schwarz (2020) look at the link between social media and hate crime.

can explain out-group hatred.

Charles Darwin (1871) has already argued that sex ratios are the primary driver of male-male competition and can explain behavioral differences between men and women.[9] Male-male competition and sexual selection might also explain why men display more out-group hatred and xenophobic attitudes than women, especially when the out-group is composed of other men (McDonald, Navarrete, and Van Vugt 2012; Buss 1988; McDonald, Navarrete, and Sidanius 2011; Vugt, Cremer, and Dirk P. Janssen 2007b).[10] Concerns over the relationship between extreme sex ratios and violent crime have been raised previously.[11] Several studies have shown that a surplus of males is associated with more aggression and violence (see e.g. Hesketh and Xing 2006; Edlund et al. 2013; Cameron, Meng, and Zhang 2017). Unmarried men are more likely to engage in crime and intimate partner violence (Abramsky et al. 2011; Sampson, Laub, and Wimer 2006). More recently, Baranov, De Haas, and Grosjean (2021) have shown that skewed gender ratios can shape behavioral norms for men

9. See also the work of other evolutionary biologists (e.g. Bachtrog et al. 2014; Sapolsky 1990, 1991; Emlen and Oring 1977) and psychologists (Buss 1994, 1988). Hesketh and Xing (2006) propose that especially men of low economic status struggle to find a partner when the competition for women is high and therefore suffer from marginalization.

10. The male warrior hypothesis states that the greater variance in terms of reproductive success among men as compared to women and their lower level of obligatory parental investment have enhanced intra-sexual competition in men and can explain differences in physical aggressiveness between the sexes (Vugt, Cremer, and Dirk P. Janssen 2007b).

11. When women are relatively scarce and the competition among men intensifies this also has economic consequences as it affects how men and women interact with each other within the household. One consequence is that female labor force participation increases (see e.g. Teso 2018; Goldin and Olivetti 2013; Heer and Grossbard-Shechtman 1981; Chiappori, Fortin, and Lacroix 2002; Amuedo-Dorantes and Grossbard 2007; Grossbard 2015) and in the long run this has been shown to affect norms about female work more generally (Grosjean and Khattar 2018). This is also in line with the behavior predicted by marriage market models (Becker 1973, 1974). Men, on the other hand, adjust their behavior in a way such that they become more attractive partners to females (see e.g. Wei and Zhang 2011; Guttentag and Secord 1983; Pedersen 1991). Marriage rates increase and men engage in more committed relationships when women are relatively scarce (Grosjean and Khattar 2018; Schacht and Kramer 2016). The quality of marriage increases and they get more invested as fathers (Otterbein 1965; Grosjean and Brooks 2017; Schmitt 2005).

more generally.[12] In this study, we focus on male-male competition as a potential explanation for hatred towards *foreigners*. Our study is therefore closest to Dancygier et al. (2022) who also study the case of Germany and provide evidence on the relationship between sex ratios and hatred towards foreigners.[13] We extend their analysis in at least two ways. First of all, using an instrumental variables strategy we aim at providing causal evidence for the relationship between sex ratios and anti-refugee hate crime. Second, by adding the vote share of the right-wing AfD as an outcome we show that skewed sex ratios do not only affect extreme forms of behavior such as hate crime but also political attitudes more generally.

Lastly, our study adds to a growing literature studying the "decline of men" and the "crisis of masculinity". Several studies have shown that women have overtaken men in education (Becker, Hubbard, and Murphy 2010). Recent developments in the labor market, mostly the change from a focus on the traditionally more male manufacturing sector to the traditionally more female service sector have further contributed to the worsening of the relative position of men (see e.g. Autor, Dorn, and Hanson 2019).[14] In addition, men often adhere to traditional gender norms which prevents them from taking on "female" jobs (George A. Akerlof and Rachel E. Kranton 2000b, 2010) and prefer to stay unemployed instead (Katz 2014; Delfino 2021). A cultural change towards more gender equality furthermore constitutes a threat to male identity if

12. They show that masculinity norms are more pronounced in areas that were historically male-dominated. As a consequence, men engage more in violent behavior and they also observe, among other outcomes, more male-stereotypical occupational segregation and a lower support for same-sex marriages.

13. They use survey measures to confirm that individual-level support for hate crime increases when men anticipate that the arrival of refugees lowers their own chances to be successful in the partner market.

14. Cortes, Jaimovich, and Siu (2021) provide evidence that women rather than men have sorted into high-paying occupations when social tasks have started to become more important for those.

that is built on a more traditional idea of masculinity. Kimmel (2013) and Inglehart and Norris (2019) discuss how this has transformed the political landscape in the US and in Europe. By studying the partner market as a source for the anger of young men and the potential social unrest it creates, this study adds a new dimension to the debate on the status of young men in society.

The next section provides an overview of the demographic conditions and the institutional setup in Germany. Section 3.2.1 describes the situation in East Germany and Section 3.2.2 gives a short overview of the events leading to a rise in right-wing extremism in Germany in recent years. Section 3.3 introduces and discusses the data sources used for this study. In Section 3.3.3 we present our empirical strategy, introduce our instrument and discuss the results.

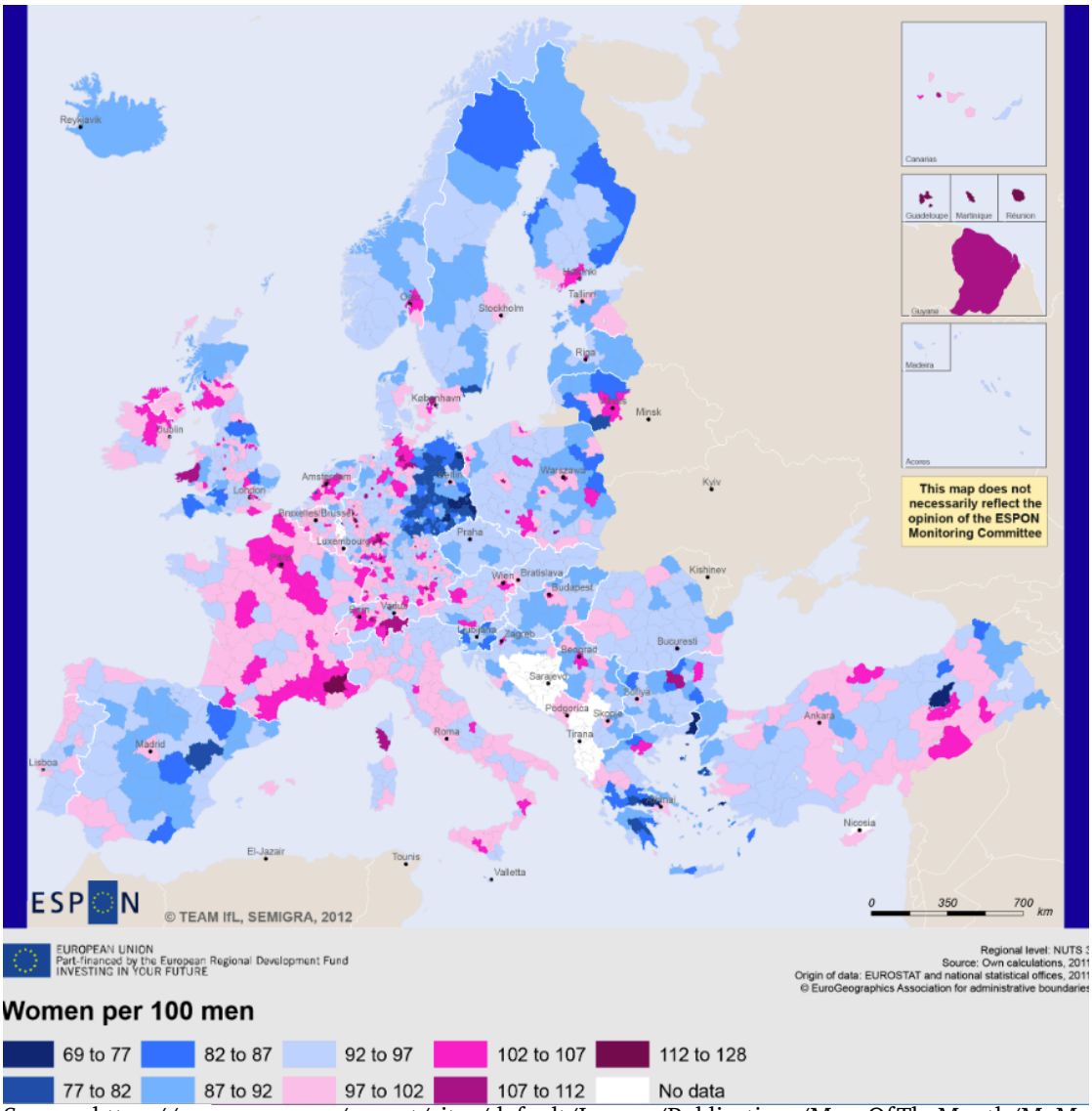
3.2 Background

3.2.1 The Lack of Women in East Germany

Figure 3.1 shows the number of women per 100 men in the age group 25 to 29 in the year 2008 in Europe. The figure makes regional differences in gender ratios strikingly apparent and highlights the exceptional position of East Germany. No other contiguous region in Europe displays a comparable lack of women. The share of women in the young population is lower than in some of the least populated parts in Europe such as the north of Norway. In the majority of districts in East Germany there are between 69 and 87 women per 100 men. The only exception is Berlin where the gender ratio is balanced with 87 to 102 women per 100 men.

The lack of women in East Germany is largely caused by gender specific migration. Women are more likely than men to migrate from East to West Germany and they

Figure 3.1: Number of Women per 100 Men in the Age Group 25 to 29 in 2008



Source: https://www.espon.eu/export/sites/default/Images/Publications/MapsOfTheMonth/MoM_February2013/SEMIGRA_High-Quality.png, accessed 20 May 2022.

are also less likely to return to the East which prevents sex ratios from equalizing for those groups who are in the age to form a family (Leibert 2016). Most of the available studies have looked at the initial migration flows from East to West Germany covering the time period between the Fall of the Berlin Wall in 1989 and the early 2000s. The

results indicate that migration from East to West Germany is indeed mostly driven by young people (Hunt 2006). Young individuals leaving the East tend to be positively selected in terms of education. This effect, however, seems to be driven mostly by men, at least for the period under study. Fuchs-Schündeln and Schündeln (2009) find that women go West to take up jobs which require a low skill level and Melzer (2011) finds that females with intermediate secondary education are less mobile than those in the group with the lowest level of education.

In a more recent study, Stauder (2018) looks at migration up to 2012 and shows that young women move West to seek further education rather than for job-related reasons. The same is true for young men moving from West to East Germany. Apart from gender disparities in educational attainment, the different labor market structures in East and West Germany have been considered a reason for why women decide to migrate West. The East German labor market with its focus on the manufacturing sector favors men, whereas women see better opportunities in the West which offers more jobs in the service sector (see e.g. Kröhnert and Klingholz (2006) and also Black and Spitz-Oener (2010) for the effect of technological change on female employment and wages).

These migration patterns clearly have an impact on the partner market. Eckhard and Stauder (2018) find that mating chances for men have strongly deteriorated in East Germany as a result of gender-selective migration. At the same time, the partner market is potentially yet another factor for why more women than men move to the West. Fuchs-Schündeln and Schündeln (2009) find that more unmarried women than unmarried men leave East Germany towards the West. This is in line with the suggestion of Kröhnert and Klingholz (2006) that women move to West Germany to

find partners with similar levels of education.

3.2.2 Right-wing Extremism in Germany

In the years 2015 to 2017, almost 1.4 million refugees arrived in Germany and sought asylum. More than half of the applicants were individuals from Syria, Iraq and Afghanistan (Bundesamt für Migration und Flüchtlinge 2019). Whereas one part of the population welcomed and some even hosted refugees, another part was opposed to the decision of Angela Merkel to open the borders to people fleeing their home countries in autumn 2015. Instances of hate crime against migrants and refugees increased significantly in the aftermath. In 2015, almost 1,500 attacks on refugee homes were officially recorded - five times the number of 2014.[15] In 2016, the authorities registered almost 10 attacks on migrants per day.[16]

The demographic characteristics of the refugee population became a fundamental point in the discussion after the events on New Year's Eve 2015/2016. More than 1,200 women reported that they were sexually assaulted in that night. Many men who were later identified as suspects were asylum seekers.[17] A majority of the refugee population who arrived during that time is indeed male and many are between 15 and 40 years old (Statistisches Bundesamt 2018). Many male refugees were also unmarried and single when they arrived in Germany.[18]

For a long time, the topic dominated and polarized the public debate. It was therefore expected to affect voting outcomes in the 2017 parliamentary elections. Indeed, the Alternative for Germany (Alternative für Deutschland - AfD) was triumphant with

15. See Die Zeit (2016).
16. See BBC (2017).
17. See Die Zeit (2017).
18. For details see Bundesinstitut für Bevölkerungsforschung (2019).

its campaign against refugees and migrants. When the party was founded in 2013, it gained popularity mainly due to its Euro-sceptical program and barely missed entry into parliament the same year. Just before the 2017 elections, the party decided to shift its focus to the issues of migration and national identity. It obtained almost 13 percent of the votes and thus became the third strongest force in the German parliament, overtaking the Liberals, the Greens and the Left party. According to the 2019 Chapel Hill Expert Survey, the party is among the five parties with the most restrictive views on immigration in the universe of parties the survey covers (Jolly et al. 2022). Since March 2022, the party is under surveillance by the Federal Office for the Protection of the Constitution.[19]

3.3 Missing Women and Right-wing Attitudes

3.3.1 Data

3.3.1.1 Demographics and Voting Outcomes

For our analysis, we rely on mainly two data sources. First, we use data from the data base INKAR which is provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development.[20] In the INKAR data base, the institute collects indicators on the demographic conditions as well as, for example, information on labor market conditions and voting outcomes on the level of various spatial units. We use data on the NUTS-3 level which is the level of districts or "Kreise" in Germany. There are in total 401 districts in Germany. We use information for the year 2017 with the exception of the gender ratio. The gender ratio is defined as the number of

19. Figure 3.A.1 in the Appendix shows a selection of posters the AfD used during its 2017 electoral campaign, demonstrating the programmatic focus of the party.
20. See https://www.inkar.de/.

women divided by the number of men in the population. We focus on women and men between 20 and 40 years old as we are interested in the effect of a skewed gender ratio when adult individuals are most active in the partner market. We decided to use the gender ratio of the 20 to 40 year-olds as of 2014, i.e. before the large influx of refugees starting at the end of 2015, in our analysis. As many refugees were young males this created even more skewed gender ratios in many regions. However, we are interested in regions where the partner market was already tight for men.

3.3.1.2 Violence against Refugees

The second main data source we are using is a collection of anti-refugee incidents provided by the Amadeu Antonio Foundation and Pro Asyl for the year 2015 to 2017.[21] The cases are classified by type of incidents and are either arson attacks against refugee homes, other forms of attacks against refugee homes, any form of physical attacks against refugees and protests against refugees. We summarize them under the term hate crime. In our analysis, we include the first three types and exclude cases which were still under investigation. The data also provides information on the place of the attack and its latitude and longitude. We use that information to assign places to districts. We sum the number of cases over the entire period for each district and normalize these numbers using the mean of the population counts in the districts over the same three years.[22]

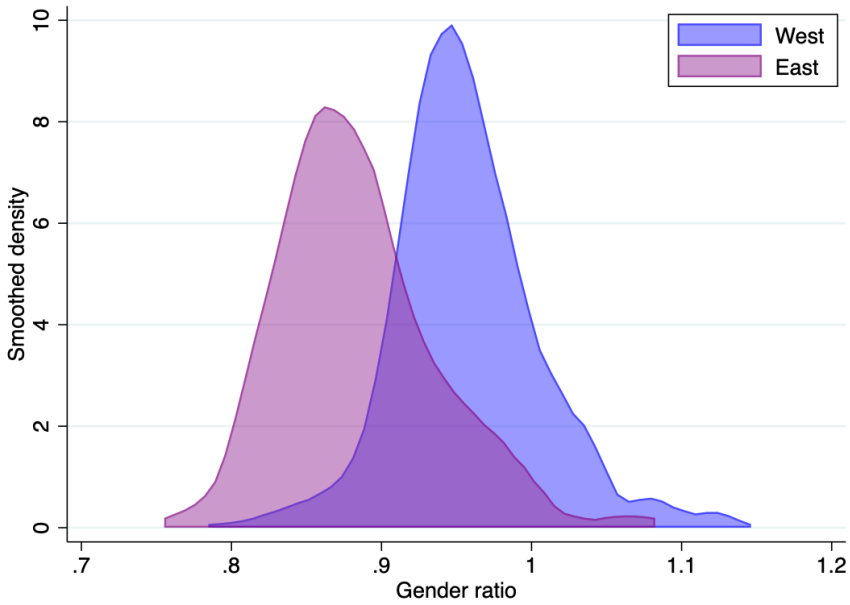
21. The data can be found at https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle.
22. For discussion of data quality see Müller and Schwarz (2020).

3.3.2 Descriptive Analysis

3.3.2.1 Gender Ratios and Hate Crime

Figure 3.2 shows the distribution of the gender ratio for people between 20 and 40 years of age in 401 districts in the year 2014, separately for East and for West German districts. The distribution in West Germany is shown in blue and the distribution in the East is shown in purple. This close-up on gender ratios within Germany shows that there is a clear divide between East and West Germany. In West German districts gender ratios among the younger generations are much higher than in the West. The median gender ratio in West German districts is 0.93 and in the East it is only 0.87. The map in Panel (a) of Figure 3.3 illustrates a closeup of Germany of the

Figure 3.2: Distribution of Gender Ratios in East and West German Districts in 2014



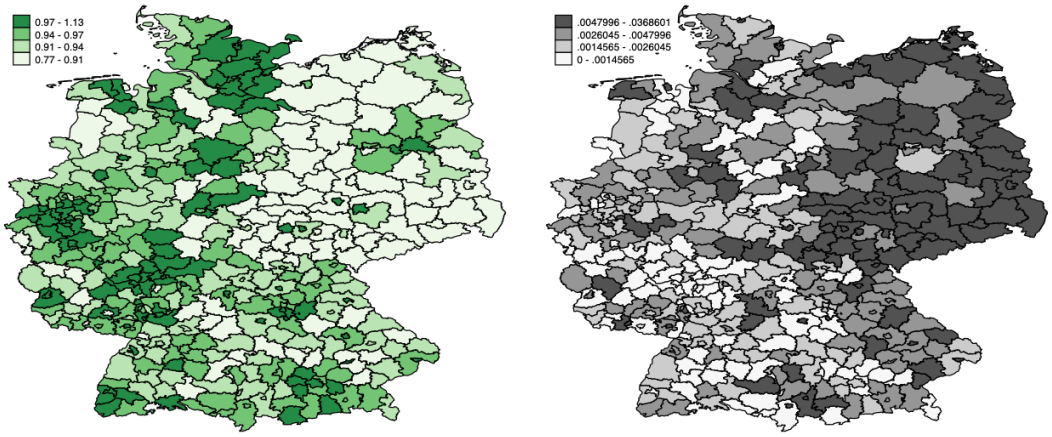
Note: The graph shows the distribution of gender ratios for 401 East and West German districts, in purples and blue respectively, for the year 2014. The gender ratio is defined as the number women divided by the number of men. Only individuals between the age of 20 and 40 are considered. The data was obtained from the INKAR database (https://www.inkar.de/).

map shown in Figure 3.1. We see that the majority of East German districts is in the category with the lowest gender ratio between 0.77 and 0.91, i.e. they have an extreme surplus of men. With a gender ratio between 0.94 and 0.97 the gender ratio is only mildly better in districts bordering the capital. Only Berlin itself and some other larger cities such as Leipzig, Erfurt, Potsdam or Weimar have a more balanced gender ratio. The map also shows that not only East German districts are affected by extreme sex ratios. Especially in rural areas, like for example in Bavaria and Baden-Wuerttemberg, we see an equally large surplus of men whereas more industrial and urban areas are more like to have a surplus of women. Panel (b) of Figure 3.3 shows the number of hate crimes committed in the different districts between the years 2015 and 2017 normalized by population size. The map reveals that the number is particularly high for East Germany where almost all districts are either in the category with the highest or with the second highest share of hate crimes. No other contingent region in Germany displays a comparably high occurrence of attacks on refugees as the East.

3.3.2.2 Gender Ratios and Voting for the Alternative für Deutschland

Panel (b) of Figure 3.4 shows the share of second votes obtained by the AfD in the 2017 federal elections. The popularity of the party in the East German districts is striking. In almost all districts, more than 16.5 percent of the voters wanted to the see the party represented in the German Parliament. In four districts, a third of the votes were given to the AfD. In three voting districts, the party even won the direct mandate for which the majority of first votes is required. Prior to the national elections, the party had also been very successful in the state-level elections in East

Figure 3.3: Gender Ratios and Hate Crime in Germany



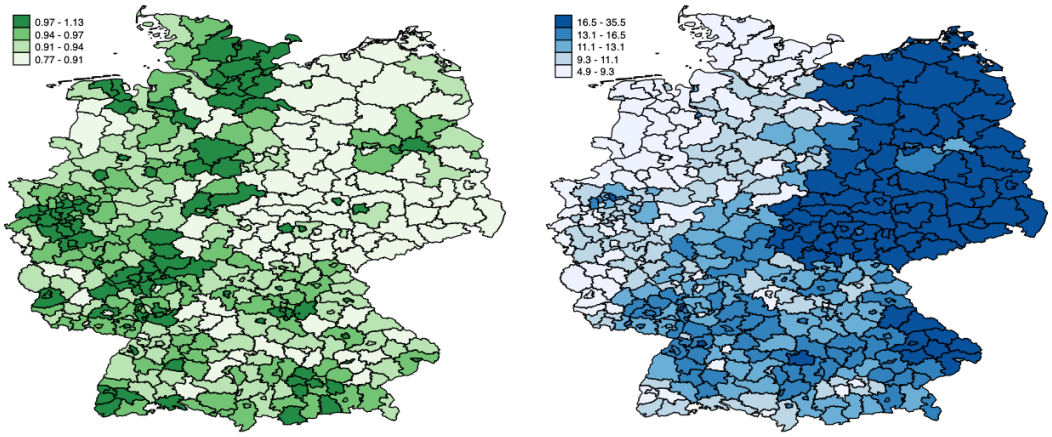
(a) Gender ratio **(b)** Hate Crime

Note: Panel (a) of Figure 3.3 shows the gender ratio in the 401 German districts in 2014 grouped into quartiles, i.e. the different shades of green correspond to one of four quartiles as described on the left. The gender ratio is defined as the number of women divided by the number of men in a district. Only individuals between 20 and 40 years were considered. The data was obtained from the INKAR database (https://www.inkar.de/). Panel (b) shows the incidence of hate crime for the districts. The data for attacks on refugees was obtained from the project *Mut gegen rechte Gewalt* (https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle). The above shows the number of incidents for the years 2015 to 2017 for each district divided by the mean of the population sizes for the same three years. The different shades of gray correspond to one of four quartiles as described on the left.

Germany.[23] The comparison with the map of gender ratios in Panel (a) reveals a striking correspondence between a surplus of men and the success of the AfD. This is not only true for East Germany but also for other parts of the country. For example, we see that the party was also very successful in the Czech border region of Bavaria where we observe a similarly low share of women as in the East.

23. Since 2014, the party had been successful in all state-level elections prior to the national elections in 2017.

Figure 3.4: Gender Ratios and AfD Votes



(a) Gender ratio **(b)** AfD Votes

Note: Panel (a) of Figure 3.3 shows the gender ratio in the 401 German districts in 2014 grouped into quartiles, i.e. the different shades of green correspond to one of four quartiles as described on the left. The gender ratio is defined as the number of women divided by the number of men in a district. Only individuals between 20 and 40 years were considered. The data was obtained from the INKAR database (https://www.inkar.de/). Panel (b) shows the vote share obtained by the Alternative für Deutschland in the 2017 national elections. The votes shown are the second votes which determine the final share of seats a party obtains in parliament. The data was obtained from the INKAR database (https://www.inkar.de/) and the numbers are grouped into quintiles as shown on the left of the map.

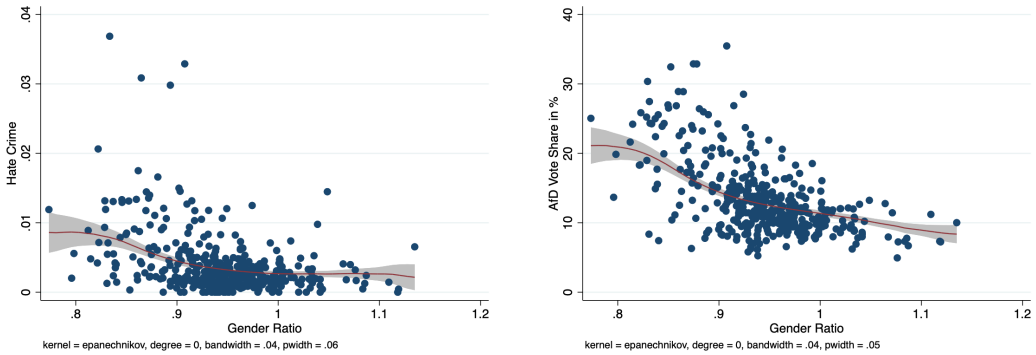
3.3.3 Gender Ratios and Xenophobia

3.3.4 Empirical Strategy

As the descriptive analysis in the previous section shows, both hate crime and the support for the Alternative für Deutschland are negatively related to the gender ratio of the 20 to 40-year-olds in a district. More specifically, a higher share of women in the local population is associated with lower levels of hate crime and a lower support for the AfD. Figure 3.5 shows that both the vote share obtained by the AfD and the

share of hate crimes committed drops quite substantially - by about 50 percent - in regions where the gender ratio is more balanced, i.e. where the ratio of women to men is 0.95 or higher.

Figure 3.5: Local Polynomial Smoothing



(a) Hate Crime **(b)** AfD Votes

Note: Panel (a) and Panel (b) show the correspondence between the gender ratio in German districts for 20 to 40 year olds in 2014 and hate crime and votes obtained by the Alternative für Deutschland in the 2017 national elections, respectively, both in percent. The data for attacks on refugees was obtained from the project *Mut gegen rechte Gewalt* (https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle). The above shows the number of incidents for the years 2015 to 2017 for each district divided by the mean of the population sizes for the same three years. Data on vote shares was obtained from the INKAR database (https://www.inkar.de/). The red lines are the results from local polynomial smoothing. The gray-shaded area represents the 95% confidence interval.

This is confirmed by the results from a simple regression of the following form:

$$y_i = \alpha + \beta_s ratio_{i,2014} + x_i'\gamma + \varepsilon_i. \tag{3.1}$$

In the above equation, y_i is the outcome of interest, i.e. usually the vote share obtained by the AfD in the 2017 national elections or the share of hate crimes committed in a district between 2015 and 2017 normalized by the population count. The gender ratio for individuals between 20 and 40 years old in each district i in 2014 is described by $ratio_{i,2014}$. We also include controls on the district level (x_i).

The error term is given by ε_i. Table 3.1 depicts the results from this regression. In all specifications we see the strong negative relationship between gender ratios in 2014 and voting outcomes and hate crime, respectively, which is significant on a one-percent significance level. The results suggest that an increase of the gender ratio by 1 percentage point lowers hate crime by about 0.002 to 0.0003 percentage points and decreases the vote share obtained by the AfD by 0.46 to 0.54 percentage points depending on whether we control for regional factors or not.

We control for several factors that we expect to be highly relevant for xenophobic sentiments and voting behavior at the regional level. First of all, we control for population density to take any differences between rural and urban areas into account. This indeed seems to be a relevant factor as we see a significant coefficient for this control variable. However, it is only of relatively small magnitude. This is also true when we control for the share of foreigners. This is not too surprising as asylum seekers were usually assigned to districts in proportion to the pre-existing population. Especially, in East Germany where we see higher vote shares for the AfD and a higher incidence of hate crime the share of foreigners and asylum seekers is typically rather low.[24] We also find a significantly positive coefficient when we control for the share of unemployed as a measure for local economic conditions, i.e. as expected higher unemployment rates are associated with more xenophobic sentiments in the population. The industry structure of a region, proxied by employment in the different sectors, does not seem to be a relevant factor above and beyond the level of unemployment.

24. See for example https://wahlatlas.net/euw/19/#!Schutzsuchende2017jeTsdEw,default, accessed 20 May 2022.

Table 3.1: Regression Results

	(1) Hate Crime	(2) AFD Votes	(3) Hate Crime	(4) AfD Votes	(5) Hate Crime	(6) AfD Votes
Gender Ratio	-0.030***	-54.282***	-0.019***	-47.445***	-0.022***	-46.346***
	[0.005]	[4.359]	[0.005]	[4.558]	[0.007]	[5.445]
Population density			0.000*	0.001***	0.000*	0.001***
			[0.000]	[0.000]	[0.000]	[0.000]
Share of foreigners			-0.000***	-0.303***	-0.000***	-0.294***
			[0.000]	[0.069]	[0.000]	[0.074]
Share unemployed			0.000***	0.225**	0.000**	0.292***
			[0.000]	[0.099]	[0.000]	[0.106]
Employment in secondary sector					-0.000	-0.083
					[0.000]	[0.206]
Employment in tertiary sector					-0.000	-0.111
					[0.000]	[0.214]
Constant	0.032***	64.588***	0.023***	59.300***	0.051**	67.816***
	[0.005]	[4.227]	[0.004]	[4.391]	[0.023]	[19.839]
Mean dep. var.	0.004	13.389	0.004	13.389	0.004	13.354
Obs.	401	401	401	401	393	393

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table shows the results from a regression based on Equation 3.1. The outcome of interest is either hate crime or votes obtained by the AfD (see specification in the table head). Data on hate crime was obtained by "Mut gegen rechte Gewalt". Incidents of refugee attacks were summed up on the level of districts for the years 2015 to 2017 and normalized by population size. The remaining information was obtained from the INKAR database (https://www.inkar.de/). All variables are in percent and with the exception of population density which is defined as population per square kilometer. Apart from the gender ratio and the hate crime variable all variables are for the year 2017. Specification (5) and (6): Due to missing values in some districts for employment shares in the different sectors the sample size drops to 393 districts. Employment in the primary sector is used as the reference category.

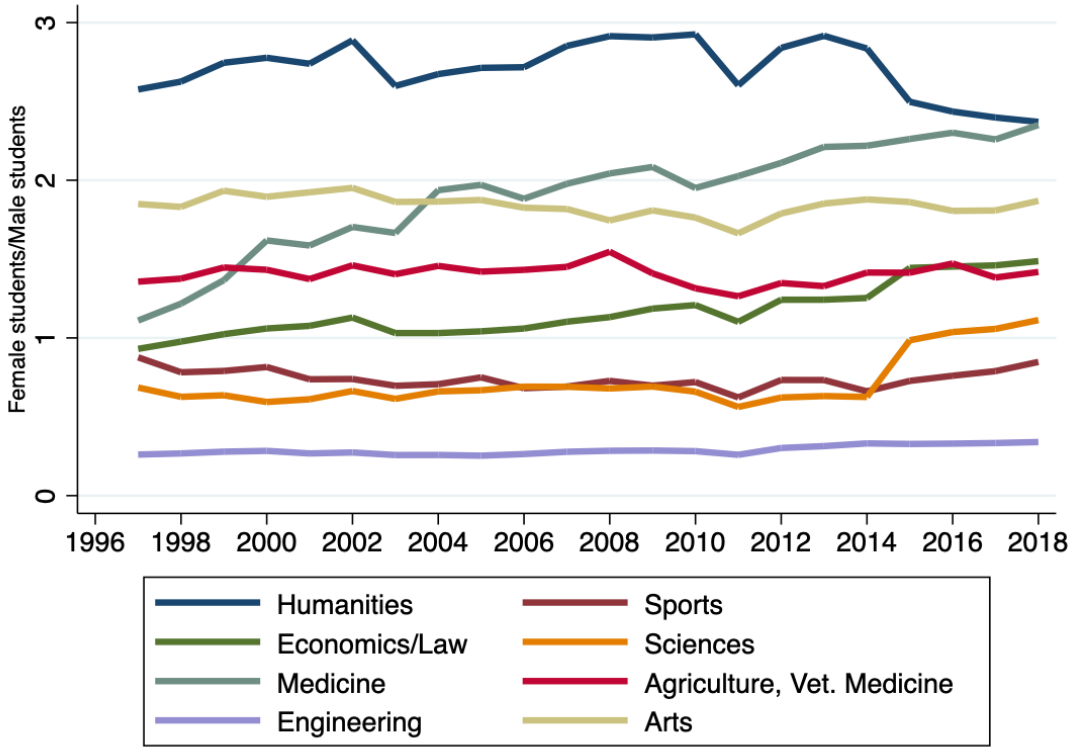
3.3.4.1 The Instrument

The number of women living in an area is likely to be related to factors that are determinants of xenophobic attitudes and voting behavior as well. Section 3.2.1 discusses for example the importance of labor market conditions for individuals' decision to migrate internally and it has been shown that these are also relevant for voting.[25] It could also be that people living in a region share certain values which affect not only how they vote but also make a community more or less attractive to live in for one of the two sexes. The gender ratio in Equation 3.1 is therefore not exogenous and the initial results from a simple regression are likely to be biased.

We therefore instrument for the gender ratio using a variable which describes the share of student places usually taken up by women in each district and show that this is positively related to the share of young women living there. To construct the measure of female student places we use data from the Federal Statistical Office of Germany.[26] First, we use information on the share of student places taken up by women in the different subject groups. Figure 3.6 illustrates the data. It shows the ratio of female to male students enrolled in seven different subject groups for the years 1998 to 2018. In Germany, the share of women in the STEM subjects is traditionally low. The Figure shows that this is particularly true for engineering. In other - traditionally more male subjects - such as science, sports or economics and law there has been a small trend towards an increase in the share of females in more recent years but the number of women deciding to study engineering remains remarkably low. Women are more likely to study humanities and increasingly also

25. See e.g. Falk, Kuhn, and Zweimüller (2011).
26. See Destatis, Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Fächergruppen, Hochschularten und Geschlecht, Table 2.5.21, 1997 to 2018.

Figure 3.6: Gender Ratio in Subject Groups at German Universities, 1997 to 2018



Note: The graph shows the ratio of female to male students enrolled at German universities and universities of applied sciences in eight different subject groups between 1917 and 2018. The data was obtained from Destatis (Publication: Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Fächergruppen, Hochschularten und Geschlecht, Table 2.5.21, 1997 to 2018).

take up studies in medicine. Overall, the graph clearly demonstrates that men and women decide to study different subjects. The gender composition of a university will therefore depend on the subjects it offers.

In a second step, we combine this information with data on the number of students by subject group enrolled at all German universities. We geolocate universities and determine for each district all universities which are within a 50km radius from the centroid of the district. Using the share of women typically enrolled in each subject as presented in Figure 3.6, we have then compute a measure for the sum of women

enrolled at all universities within that radius by multiplying the students in each subject group with the share of women typically enrolled in that subject. We then sum over universities and divide the number of women enrolled according to our measure by the number of all students. Lastly, we compute for each district a weighted average of its own share of female students and those of the neighboring districts using inverse distances between the district centroids as weights. This is meant to create a more natural measure and reflects that district borders are arbitrary in this context.

Figure 3.7 shows the share of student places typically taken up by women in 2014, the year in which we also measure the gender ratio. Especially in East Germany there are several districts where more than 50 percent of student places go to female students as depicted in Figure 3.8. The graph shows the distribution of our measure of female student places separately for East and for West Germany. The share of student places taken up by women is indeed highly correlated with the gender ratio of 20 to 40 year olds as is also shown in the scatter plot in Figure 3.9. There is a strong and significantly positive relationship between the two which allows us to use the share of student places typically taken up by women in 2014 as an instrument for the gender ratio of 20 to 40 year olds in the same year.[27] For the exclusion restriction to hold one would require that the share of student places typically held by females only affects voting outcomes through the effect it has on gender ratios.

3.3.4.2 Results

Table 3.2 shows the results from our instrumental variable strategy. We find that the negative relationship between gender ratios and hate crime respectively voting for the AfD persists and continues to be significant. Table 3.2 shows the IV results.

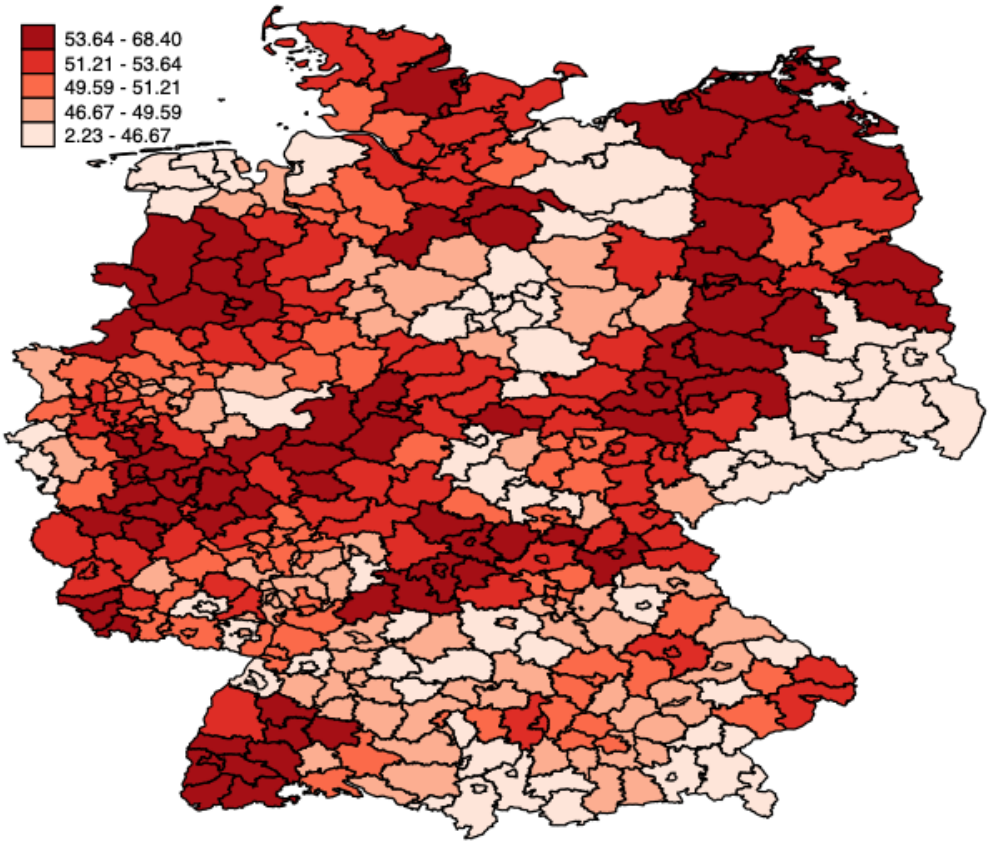
27. The F-statistic from a simple regression of the gender ratio on our instrument is 13.24.

The coefficients are now larger in absolute values. Whereas the coefficient on the effect of gender ratios on AfD votes continues to be significant on a one percent level, significance has reduced to the 5 percent level for the coefficient indicating its effect on hate crime. However, all coefficients are robust to the inclusion of controls which show a similar pattern as in the basic regression results.[28] According to our IV results one more women per 100 men is associated with a decrease in the share of hate crimes by 0.6 to 0.9 percentage points and a decrease in votes for the Alternative in Germany by 0.71 to 0.93 percentage points, depending on the specification.

An alternative explanation for our results could be that the violent behavior against and the hatred towards refugees is just one expression of violence rather than violence specifically directed at one group of people. As Hesketh and Xing (2006), Edlund et al. (2013) and Cameron, Meng, and Zhang (2017) show, a surplus of males can increase aggression and violence as seen in crime rates, for example. If that was driving our results, we would expect this to also show in other forms of crime, in particular in levels of violent crime, sex crime and physical assault more generally. Using data from the Federal Criminal Police Office on the number of offenses committed in these three categories we therefore repeat the IV estimation. The results are presented in Table 3.3. We do not find any significant effects of gender ratios on crime rates. This suggests that differences in aggressive behavior alone are unlikely to drive our results and provides another motivation to better understand the link between the origins of right-wing hate crime and voting and the basic biological and social needs of humans.

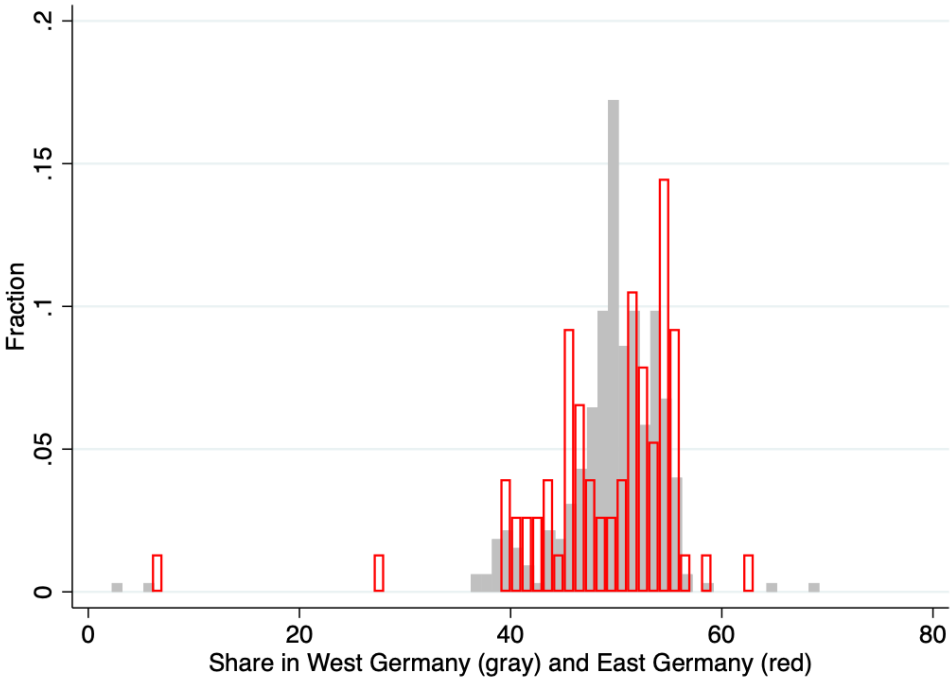
28. The only exception is that unemployment does not significantly affect our outcomes of interest anymore.

Figure 3.7: Student Places Typically Taken up by Women, 2014



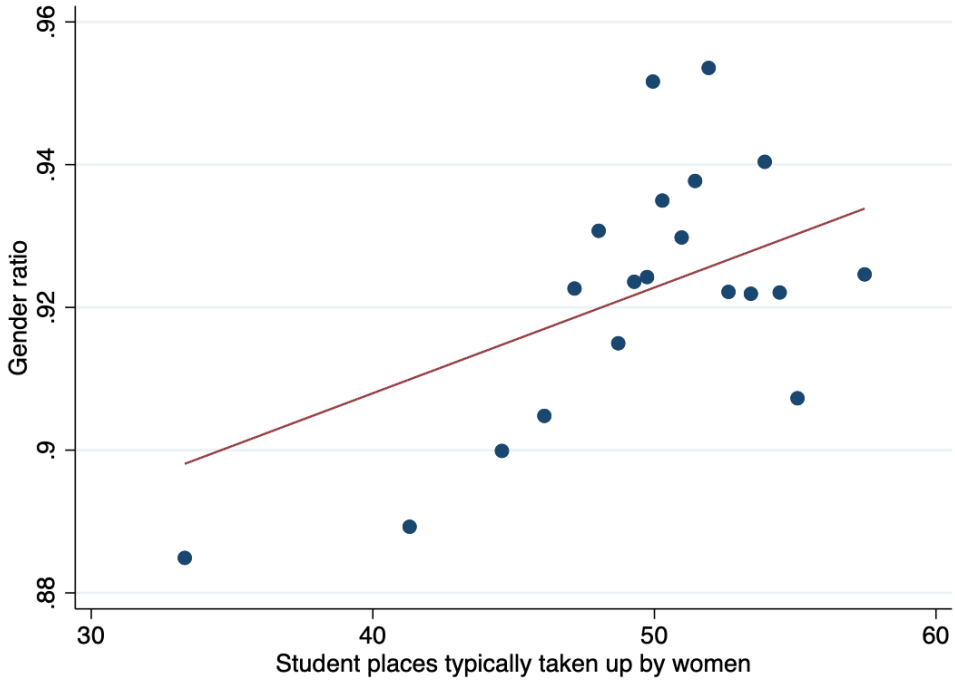
Note: The map shows the availability of student places typically taken up by women in 2014 according to our definition (see Section 3.3.4.1 for a detailed description). We have used information obtained from Destatis on the number of students enrolled in all German universities and used information on the share of student places typically taken up by women in the different subject groups (Destatis, Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Fächergruppen, Hochschularten und Geschlecht, Table 2.5.21, 1997 to 2018) to compute a measure of the share of student places (in percent) typically taken up by women. The map groups the share into five quintiles represented by different shades of red as described on the left.

Figure 3.8: Distributions of Student Places Typically Taken up by Females in 2014



Note: The graph shows the distribution of our measure for the availability of student places typically taken up by women in 2014 (see Section 3.3.4.1 for a detailed description), separately for East Germany (in red) and West Germany (in gray). We have used information obtained from Destatis on the number of students enrolled in all German universities and used information on the share of student places typically taken up by women in the different subject groups (Destatis, Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Fächergruppen, Hochschularten und Geschlecht, Table 2.5.21, 1997 to 2018) to compute a measure of the share of student places (in percent) typically taken up by women.

Figure 3.9: Relationship between Female Study Places and Gender Ratios in 2014



Note: The graph shows the relationship between our measure for the availability of student places typically taken up by women in 2014 (see Section 3.3.4.1 for a detailed description) and the gender ratio for 20- to 40-year olds in the German districts in 2014. We have used information obtained from Destatis on the number of students enrolled in all German universities and used information on the share of student places typically taken up by women in the different subject groups (Destatis, Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Fächergruppen, Hochschularten und Geschlecht, Table 2.5.21, 1997 to 2018) to compute a measure of the share of student places (in percent) typically taken up by women. Information on gender ratios was obtained from the INKAR database (https://www.inkar.de/). The red line presents the prediction from a linear regression of gender ratios on female student places. The slope coefficient has a value of 0.0017 and is significant on the 1% level (standard error 0.0005).

Table 3.2: IV Results

	(1) Hate Crime	(2) AFD Votes	(3) Hate Crime	(4) AfD Votes	(5) Hate Crime	(6) AfD Votes
Gender Ratio	-0.059**	-71.464***	-0.071**	-82.755***	-0.087**	-92.529***
	[0.027]	[20.469]	[0.030]	[23.133]	[0.039]	[28.129]
Population density			0.000**	0.002***	0.000**	0.002***
			[0.000]	[0.001]	[0.000]	[0.001]
Share of foreigners			-0.000***	-0.230***	-0.000**	-0.220**
			[0.000]	[0.082]	[0.000]	[0.087]
Share unemployed			0.000	0.001	-0.000	-0.103
			[0.000]	[0.172]	[0.000]	[0.249]
Employment in secondary sector					0.000	0.167
					[0.000]	[0.289]
Employment in tertiary sector					0.000	0.244
					[0.000]	[0.334]
Constant	0.059**	80.795***	0.071**	92.454***	0.069**	80.689***
	[0.025]	[19.314]	[0.028]	[21.720]	[0.028]	[24.168]
Mean dep. var.	0.004	13.389	0.004	13.389	0.004	13.354
Obs.	401	401	401	401	393	393

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table shows the results from a regression based on Equation 3.1 where we use our measure for the share of student places typically taken up by women in 2014 as an instrument for the gender ratio. For details on the construction of the instrument and the data used see Section 3.3.4.1. The outcome of interest is either hate crime or votes obtained by the AfD (see specification in the table head). Data on hate crime was obtained by "Mut gegen rechte Gewalt". Incidents of refugee attacks were summed up on the level of districts for the years 2015 to 2017 and normalized by population size. The remaining information was obtained from the INKAR database (https://www.inkar.de/). All variables are in percent with the exception of population density which is defined as population per square kilometer. Apart from the gender ratio and the hate crime variable all variables are for the year 2017. Specification (5) and (6): Due to missing values in some districts for employment shares in the different sectors the sample size drops to 393 districts. Employment in the primary sector is used as the reference category.

Table 3.3: IV Results - Other Crime

	(1) Violent Crime	(2) Sex Crime	(3) Physical Violence	(4) Violent Crime	(5) Sex Crime	(6) Physical Violence	(7) Violent Crime	(8) Sex Crime	(9) Physical Violence
Gender Ratio	0.004	0.000	0.001	0.001	0.000	-0.002	-0.001	0.000	-0.004
	[0.004]	[0.000]	[0.003]	[0.002]	[0.000]	[0.002]	[0.003]	[0.000]	[0.002]
Population density				0.000***	0.000**	0.000***	0.000***	0.000**	0.000***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Share of foreigners				0.000***	0.000	0.000***	0.000**	-0.000	0.000**
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Share unemployed				0.000***	0.000***	0.000***	0.000***	0.000**	0.000***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Employment in secondary sector							0.000***	0.000	0.000***
							[0.000]	[0.000]	[0.000]
Employment in tertiary sector							0.000***	0.000	0.000***
							[0.000]	[0.000]	[0.000]
Constant	-0.002	-0.000	0.001	-0.000	-0.000	0.002	-0.009***	-0.000	-0.005**
	[0.004]	[0.000]	[0.003]	[0.002]	[0.000]	[0.002]	[0.003]	[0.000]	[0.002]
Mean dep. var.	0.002	0.000	0.002	0.002	0.000	0.002	0.002	0.000	0.002
Obs.	401	401	401	401	401	401	393	393	393

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table shows the results from a regression based on Equation 3.1 where we use our measure for the share of student places typically taken up by women in 2014 as an instrument for the gender ratio. For details on the construction of the instrument and the data used see Section 3.3.4.1. We use information on criminal offenses from the Federal Criminal Police Office (https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/PolizeilicheKriminalstatistik/2017/BKATabellen/FaelleLaenderKreiseStaedte/BKA-LKS-F-03-T01-Kreise_excel.xlsx?__blob=publicationFile&v=3) and normalize the number of cases in each category with the population size. The remaining information was obtained from the INKAR database (https://www.inkar.de/). All variables are in percent with the exception of population density which is defined as population per square kilometer. Apart from the gender ratio and the hate crime variable all variables are for the year 2017. Specification (5) and (6): Due to missing values in some districts for employment shares in the different sectors the sample size drops to 393 districts. Employment in the primary sector is used as the reference category.

3.4 Conclusion

In this study, we ask whether gender ratios are associated with xenophobic attitudes. We look at the German setting where gender ratios are skewed and many regions have a surplus of men. This is particularly severe in East Germany where there are as few as 7 women per 10 men in some regions. After a large influx of refugees at the end of 2015, Germany saw a striking increase in anti-refugee hate crime and the new extreme right-wing party Alternative für Deutschland comfortably passed the hurdle to enter the national parliament in 2017. We hypothesize that anti-refugee sentiments are associated with male-male competition in the partner market. The arrival of more than one million refugees from the Middle East, of which many where young unmarried men, might have been perceived as a threat by men who were already facing a tight partner market. Using district level data, we first show that there is a correlation between gender ratios and hate crime and vote shares obtained by the AfD, respectively. We then use an instrumental variable approach where we instrument for gender ratios with the share of student places typically taken up by women to confirm those results and show that the effect is arguably causal.

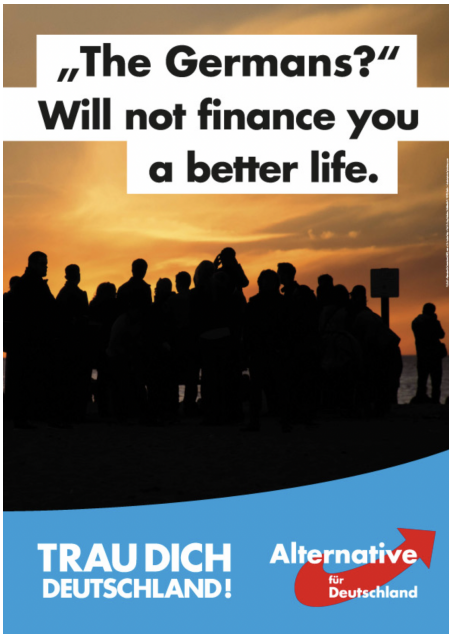
This study provides a starting point for a more detailed analysis of the phenomenon. Future work should begin by consolidating the results from the instrumental variable analysis. A potential issue is that the share of student places typically held by women could itself be an endogenous variable and would hence not fulfill the required exclusion restriction. The results should therefore be confirmed, for example, with alternative instrumental variable strategies. Another important step towards a better understanding of the relationship between competition in the partner market and out-group hatred would be to identify the underlying mechanisms. For instance,

these could be biological factors such as sexual frustration or social factors such as status concerns associated with the inability to find a partner and start a family.

Appendix to Chapter 3

3.A Posters published by the AfD

Figure 3.A.1: AfD Posters from the Campaign for the 2017 Germany National Elections.



(a) Poster 1

(b) Poster 2

(c) Poster 3

Note: The figure shows three posters which the Alternative für Deutschland used in their campaign for the 2017 national elections. The first poster in Panel (a) was taken from https://www.designtagebuch.de/die-plakate-zur-bundestagswahl-2017/afd-plakat-germans/, accessed 20 May 2022. The poster in Panel (b) says in English "New Germans? We make them ourselves." (taken from https://www.designtagebuch.de/die-plakate-zur-bundestagswahl-2017/afd-plakat-neue_deutsche/, accessed 20 May 2022) and poster three in Panel (c) says "Burkas? We are into bikinis." (https://www.designtagebuch.de/die-plakate-zur-bundestagswahl-2017/afd-plakat-bikini/, accessed 22 May 2022.). All poster use the party's slogan at the bottom which says "Be brave, Germany!".

Figure 3.A.2: AfD Poster from the Campaign for the 2019 European Elections

Note: The poster was used by the Alternative für Deutschland during the campaign for the elections of the European Parliament in 2019 in Berlin. At the top, the poster says in English "So that Europe does not become "Eurabia"!". At the bottom it says "Europeans vote for the AfD!". Source: https://www.tagesspiegel.de/berlin/afd-europawahlkampf-in-berlin-die-nackte-frau-und-die-boesen-turbantraeger/24214994.html, accessed 20 May 2022.

Bibliography

Abramsky, Tanya, Claudia Garcia-Moreno Charlotte H. Watts, Karen Devries, Ligia Kiss, Mary Ellsberg, Henrica AFM Jansen, and Lori Heise. 2011. "What factors are associated with recent intimate partner violence? findings from the WHO multi-country study on women's health and domestic violence." *BMC Public Health 11*.

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth. 2021. "Killer Incentives: Rivalry, Performance and Risk-Taking among German Fighter Pilots, 1939–45." *The Review of Economic Studies* (December).

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

———. 2005. "Identity and the Economics of Organizations." *Journal of Economic Perspectives* 19 (1): 9–32.

———. 2000b. "Economics and Identity*." *The Quarterly Journal of Economics* 115, no. 3 (August): 715–753.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages and Wellbeing*. Princeton University Press.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2020. "The polarization of reality." *AEA Papers and Proceedings* 110:324–28.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational mobility and preferences for redistribution." *American Economic Review* 108 (2): 521–54.

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. 2020. "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic." *Journal of Public Economics* 191:104254.

Amasino, Dianna, Davide Pace, and Joël Van der Weele. 2021. "Fair Shares and Selective Attention." *Working Paper*.

Amuedo-Dorantes, Catalina, and Shoshana Grossbard. 2007. "Cohort-level sex ratio effects on women's labor force participation." *Rev Econ Household* 5:249–278.

Athey, Susan, and Guido Imbens. 2018. *Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption*.

Autor, David, David Dorn, and Gordon Hanson. 2019. "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men." *American Economic Review: Insights* 1, no. 2 (September): 161–78.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased judgments of fairness in bargaining." *American Economic Review* 85 (5): 1337–1343.

Bachtrog, Doris, Judith E. Mank, Catherine L. Peichel, Mark Kirkpatrick, Sarah P. Otto, Tia-Lynn Ashman, Matthew W. Hahn, et al. 2014. "Sex Determination: Why So Many Ways of Doing It?" *PLOS Biology* 12, no. 7 (July): 1–13.

Baker, Andrew, David F Larcker, and Charles CY Wang. 2021. *How Much Should We Trust Staggered Difference-In-Differences Estimates?* Technical report, ECGI Working Paper. ECGI.

Bansak, Kirk, Jens Hainmueller, and Dominik Hangartner. 2016. "How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers." *Science* 354 (6309): 217–222.

Baranov, Victoria, Ralph De Haas, and Pauline A. Grosjean. 2021. "Men. Roots and consequences of masculinity norms." *UNSW Business School Research Paper*.

Barone, Guglielmo, Alessio D'Ignazio, Guido de Blasio, and Paolo Naticchioni. 2016. "Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior." *Journal of Public Economics* 136:1–13.

Barron, Kai. 2021. "Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains?" *Experimental Economics* 24 (1): 31–58.

Barron, Kai, and Christina Gravert. 2021. "Confidence and career choices: An experiment." *Scandinavian Journal of Economics*.

Barron, Kai, Robert Stüber, and Roel van Veldhuizen. 2019. *Motivated motive selection in the lying-dictator game.* Technical report. WZB Discussion Paper.

BBC. 2017. *Germany hate crime: Nearly 10 attacks a day on migrants in 2016.* [Online; accessed 20-May-2022]. https://www.bbc.co.uk/news/world-europe-39096833.

Beall, Alec T., and Jessica L. Tracy. 2020. "The Evolution of Pride and Shame." In *The Cambridge Handbook of Evolutionary Perspectives on Human Behavior,* edited by Lance Workman, Will Reader, and Jerome H.Editors Barkow, 179–193. Cambridge Handbooks in Psychology. Cambridge University Press.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–846.

———. 1974. "A Theory of Marriage: Part II." *Journal of Political Economy* 82 (2): S11–S26.

Becker, Gary S., William H. J. Hubbard, and Kevin M. Murphy. 2010. "The Market for College Graduates and the Worldwide Boom in Higher Education of Women." *American Economic Review* 100, no. 2 (May): 229–33.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96, no. 5 (December): 1652–1678.

Black, Sandra E, and Alexandra Spitz-Oener. 2010. "Explaining Women's Success: Technological Change and the Skill Content of Women's Work." *The Review of Economics and Statistics* 92, no. 1 (February): 187–194.

Bolsen, Toby, James N Druckman, and Fay Lomax Cook. 2014. "The influence of partisan motivated reasoning on public opinion." *Political Behavior* 36 (2): 235–262.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. "Identity, Beliefs, and Political Conflict." *Quarterly Journal of Economics* 136 (4): 2371–2411.

Bordalo, Pedro, Marco Tabellini, and David Yang. 2021. "Issue Salience and Political Stereotypes." *NBER Working Paper*.

Borusyak, Kirill, and Xavier Jaravel. 2017. *Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume.* Technical report, Working Paper.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. *Revisiting Event Study Designs: Robust and Efficient Estimation.* Technical report, Working Paper.

Bowles, Samuel. 2009. "Did Warfare Among Ancestral Hunter-Gatherers Affect the Evolution of Human Social Behaviors?" *Science* 324 (5932): 1293–1298.

Bowles, Samuel, and Herbert Gintis. 2005. "Prosocial Emotions." In *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions,* edited by Lawrence E. Blume and Steven N. Durlauf, 339–366. Oxford University Press, October.

Bundesamt für Migration und Flüchtlinge. 2019. *Migrationsbericht 2016/2017.* [Online; accessed 19-May-2022]. http://www.bamf.de/SharedDocs/%20Anlagen/DE/Publikationen/Migrationsberichte/%20migrationsbericht-2016-2017.pdf?__blob=publicationFile.

Bundesinstitut für Bevölkerungsforschung. 2019. *Bevölkerungsforschung Aktuell.* [Online; accessed 20-May-2022]. https://www.bib.bund.de/Publikation/2019/Bevoelkerungsforschung-Aktuell-2019-3.html?nn=1219524.

Bursztyn, Leonardo, Davide Cantoni, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2021. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." *American Economic Review: Insights* 3, no. 2 (June): 233–50.

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen. 2019. "Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education." *The Review of Economic Studies* 86 (4): 1487–1526.

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao. 2017. "Status Goods: Experimental Evidence from Platinum Credit Cards*." *The Quarterly Journal of Economics* 133, no. 3 (December): 1561–1595.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107, no. 11 (November): 3288–3319.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *The Quarterly Journal of Economics* 130, no. 3 (May): 1329–1367.

———. 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics* 9 (1): 131–153.

Buss, David M. 1988. "The Evolution of Human Intrasexual Competition: Tactics of Mate Attraction." *Journal of Personality and Social Psychology* 54 (4).

———. 1994. *The evolution of desire: Strategies of human mating*. Basic Books.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky. 2022. "Measuring the Welfare Effects of Shame and Pride." *American Economic Review* 112, no. 1 (January): 122–68.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2020. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*.

Cameron, Lisa, Xin Meng, and Dandan Zhang. 2017. "China's Sex Ratio and Crime: Behavioural Change or Financial Necessity?" *The Economic Journal*.

Cantoni, Davide, Felix Hagemeister, and Mark Westcott. 2019. *Persistence and Activation of Right-Wing Political Ideology*. Rationality and Competition Discussion Paper Series 143. CRC TRR 190 Rationality and Competition, February.

Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2019. "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement*." *The Quarterly Journal of Economics* 134, no. 2 (January): 1021–1077.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs*." *The Quarterly Journal of Economics* 134, no. 3 (May): 1405–1454.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110, no. 9 (September): 2964–96.

Chiappori, Pierre-André, Bernard Fortin, and Guy Lacroix. 2002. "Marriage Market, Divorce Legislation, and Household Labor Supply." *Journal of Political Economy* 110 (1): 37–72. Accessed April 28, 2022.

Choi, Jung-Kyoo, and Samuel Bowles. 2007. "The Coevolution of Parochial Altruism and War." *Science* 318 (5850): 636–640.

Clinton, Joshua, Jon Cohen, John Lapinski, and Marc Trussler. 2021. "Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19." *Science Advances* 7 (2): eabd7204.

Cortes, Guido Matias, Nir Jaimovich, and Henry E. Siu. 2021. "The Growing Importance of Social Tasks in High-Paying Occupations: Implications for Sorting." *Journal of Human Resources*.

Costa-Gomes, Miguel A, Steffen Huck, and Georg Weizsäcker. 2014. "Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect." *Games and Economic Behavior* 88:298–309.

Costa-Gomes, Miguel A, and Georg Weizsäcker. 2008. "Stated beliefs and play in normal-form games." *Review of Economic Studies* 75 (3): 729–762.

Coutts, Alexander. 2019. "Good news and bad news are still news: Experimental evidence on belief updating." *Experimental Economics* 22 (2): 369–395.

Dancygier, Rafaela, Naoki Egami, Amaney Jamal, and Ramona Rischke. 2022. "Hate Crimes and Gender Imbalances: Fears over Mate Competition and Violence against Refugees." *American Journal of Political Science* 66 (2): 501–515.

Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. John Murray.

Delfino, Alexia. 2021. *Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs*. IZA Discussion Papers 14083. Institute of Labor Economics (IZA), January.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. " Testing for Altruism and Social Pressure in Charitable Giving *." *The Quarterly Journal of Economics* 127, no. 1 (January): 1–56.

Dellavigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2016. "Voting to Tell Others." *The Review of Economic Studies* 84, no. 1 (October): 143–181.

Dewey, P. E. 1984. "Military Recruiting and the British Labour Force during the First World War." *The Historical Journal* 27 (1): 199–223.

Dickerson, Sally S., Tara L. Gruenewald, and Margaret E. Kemeny. 2004. "When the Social Self Is Threatened: Shame, Physiology, and Health." *Journal of Personality* 72 (6): 1191–1216.

Die Zeit. 2016. *Gewalt gegen Flüchtlinge nimmt weiter zu.* [Online; accessed 20-May-2022], May. https://www.zeit.de/politik/deutschland/2016-05/thomas-de-maiziere-hasskommentare-fluechtlingskrise-verrohung-gesellschaft?sort=recommended&page=10.

———. 2017. *Kölner Silvesternacht: Zwei Jahre und 36 Verurteilungen später.* [Online; accessed 20-May-2022], December. https://www.zeit.de/gesellschaft/zeitgeschehen/2017-12/koelner-silvesternacht-2015-sexuelle-uebergriffe-ermittlungen.

Dippel, Christian, Robert Gold, Stephan Heblich, and Rodrigo Pinto. 2021. "The Effect of Trade on Workers and Voters." *The Economic Journal* 132, no. 641 (May): 199–217.

Drobner, Christoph. 2021. "Motivated Beliefs and Anticipation of Uncertainty Resolution." *American Economic Review: Insights (forthcoming)*.

Druckman, James N, and Mary C McGrath. 2019. "The evidence for motivated reasoning in climate change preference formation." *Nature Climate Change* 9 (2): 111–119.

Druckman, James N, Erik Peterson, and Rune Slothuus. 2013. "How elite partisan polarization affects public opinion formation." *American Political Science Review* 107 (1): 57–79.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm. 2018. "Refugee Migration and Electoral Outcomes." *The Review of Economic Studies* 86, no. 5 (September): 2035–2091.

Eckhard, Jan, and Johannes Stauder. 2018. "Migration and the Partner Market: How Gender-Selective Relocations Affect Regional Mating Chances in Germany." *European Journal of Population / Revue Européenne de Démographie* 34 (1): 59–86.

Edlund, Lena, Hongbin Li, Junjian Yi, and Junsen Zhang. 2013. "Sex Ratios and Crime: Evidence from China." *The Review of Economics and Statistics* 95, no. 5 (December): 1520–1534.

Edo, Anthony, Yvonne Giesing, Jonathan Öztunc, and Panu Poutvaara. 2019. "Immigration and electoral support for the far-left and the far-right." *European Economic Review* 115:99–143.

Eil, David, and Justin M Rao. 2011. "The good news-bad news effect: asymmetric processing of objective information about yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38.

Emlen, Stephen T., and Lewis W. Oring. 1977. "Ecology, Sexual Selection, and the Evolution of Mating Systems." *Science* 197 (4300): 215–223.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social Media and Protest Participation: Evidence From Russia." *Econometrica* 88 (4): 1479–1514.

Enke, Benjamin. 2020a. "Moral values and voting." *Journal of Political Economy* 128 (10): 3679–3729.

———. 2020b. "What you see is all there is." *Quarterly Journal of Economics* 135 (3): 1363–1398.

Exley, Christine L. 2015. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies* 83, no. 2 (October): 587–628.

———. 2020. "Using Charity Performance Metrics as an Excuse Not to Give." *Management Science* 66 (2): 553–563.

Falk, Armin, Andreas Kuhn, and Josef Zweimüller. 2011. "Unemployment and Right-wing Extremist Crime." *The Scandinavian Journal of Economics* 113 (2): 260–285.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90, no. 4 (September): 980–994.

———. 2002. "Altruistic punishment in humans." *Nature* 415 (6868): 137–140.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica* 65 (4): 833–860.

Ferguson, Niall. 1998. *The Pity of War - 1914–1918.* London: The Penguin Press.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80, no. S1 (March): 298–320.

Fuchs-Schündeln, Nicola, and Matthias Schündeln. 2009. "Who stays, who goes, who returns?" *Economics of Transition and Institutional Change* 17 (4): 703–738.

Gächter, Simon, and Ernst Fehr. 1999. "Collective action as a social exchange." *Journal of Economic Behavior & Organization* 39 (4): 341–369.

Gaines, Brian J, James H Kuklinski, Paul J Quirk, Buddy Peyton, and Jay Verkuilen. 2007. "Same facts, different interpretations: Partisan motivation and opinion on Iraq." *Journal of Politics* 69 (4): 957–974.

Gentzkow, Matthew. 2016. *Polarization in 2016.* Technical report. Toulouse Network for Information Technology Working Paper.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102 (1): 33–48.

Gintis, Herbert. 2000. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206 (2): 169–179.

———. 2004. "The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions." Evolution and Altruism, *Journal of Economic Behavior & Organization* 53 (1): 57–67.

Goldin, Claudia, and Claudia Olivetti. 2013. "Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply." *American Economic Review* 103, no. 3 (May): 257–62.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics.*

Gotthard-Real, Alexander. 2017. "Desirability and information processing: An experimental study." *Economics Letters* 152:96–99.

Gregory, Adrian. 2008. "British 'War Enthusiasm' in 1914: a Reassessment." In *Evidence, History and the Great War: Historians and the Impact of 1914-18,* 1st ed., 67–85. Berghahn Books.

Grimalda, Gianluca, Andreas Pondorfer, and David P. Tracer. 2016. "Social image concerns promote cooperation more than altruistic punishment." *Nature Communications.*

Grosjean, Pauline, and Robert C. Brooks. 2017. "Persistent effect of sex ratios on relationship quality and life satisfaction." *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1729): 20160315.

Grosjean, Pauline, and Rose Khattar. 2018. "It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios." *The Review of Economic Studies* 86, no. 2 (May): 723–754.

Grossbard, Shoshana. 2015. “A Theory of Allocation of Time in Markets for Labor and Marriage: Macromodel.” In *The Marriage Motive: A Price Theory of Marriage: How Marriage Markets Affect Employment, Consumption, and Savings,* 21–32. New York, NY: Springer New York.

Gullace, Nicoletta F. 1997. “White Feathers and Wounded Men: Female Patriotism and the Memory of the Great War.” *Journal of British Studies* 36 (2): 178–206.

———. 2002. *The Blood of Our Sons": Men, Women, and the Renegotiation of British Citizenship During the Great War.* New York: Palgrave MacMillan.

Guttentag, Marcia, and Paul F Secord. 1983. *Too Many Women? The Sex Ratio Question.* SAGE Publications.

Haaland, Ingar, and Christopher Roth. 2021. “Beliefs about racial discrimination and support for pro-black policies.” *Review of Economics and Statistics,* 1–38.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2020. “Designing information provision experiments.” *CEBI Working Paper Series, Working Paper 20/20.*

Halla, Martin, Alexander F Wagner, and Josef Zweimüller. 2017. “Immigration and Voting for the Far Right.” *Journal of the European Economic Association* 15, no. 6 (March): 1341–1385.

Hallifax, Stuart. 2010. “‘Over by Christmas’: British popular opinion and the short war in 1914.” *First World War Studies* 1 (2): 103–121.

Hart, Peter J. 2010. “The White Feather Campaign: A Struggle with Masculinity During World War I.” *Inquiries Journal* 2 (02).

Heer, David M., and Amyra Grossbard-Shechtman. 1981. "The Impact of the Female Marriage Squeeze and the Contraceptive Revolution on Sex Roles and the Women's Liberation Movement in the United States, 1960 to 1975." *Journal of Marriage and Family* 43 (1): 49–65.

Hesketh, Therese, and Zhu Wei Xing. 2006. "Abnormal sex ratios in human populations: Causes and consequences." *Proceedings of the National Academy of Sciences* 103 (36): 13271–13275.

Hume, David. 1978. *Treatise of Human Nature.* 2nd edition. Edited by L.A. Selby-Bigge. London: Oxford University Press.

Hunt, Jennifer. 2006. "Staunching Emigration from East Germany: Age and the Determinants of Migration." *Journal of the European Economic Association* 4 (5): 1014–1037. Accessed May 11, 2022.

Imai, Kosuke, and In Song Kim. 2021. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29 (3): 405–415.

Inglehart, Ronald, and Pippa Norris. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism.* Cambridge Univer- sity Press.

Ingraham, Christopher. 2019. *The share of Americans not having sex has reached a record high.* Edited by washingtonpost.com. [Online; posted 29-March-2019], March. https://www.washingtonpost.com/business/2019/03/29/share-americans-not-having-sex-has-reached-record-high/.

Ivanov, Asen, Dan Levin, and Muriel Niederle. 2010. “Can relaxation of beliefs rationalize the winner’s curse?: An experimental study.” *Econometrica* 78 (4): 1435–1452.

Jolly, Seth, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2022. “Chapel Hill Expert Survey trend file, 1999–2019.” *Electoral Studies* 75:102420.

Kahan, Dan M. 2013. “Ideology, motivated reasoning, and cognitive reflection: An experimental study.” *Judgment and Decision making* 8:407–24.

Katz, Lawrence F. 2014. “America’s Jobs Challenges and the Continuing Role of the U.S. Department of Labor.” *ILR Review* 67:578–583.

Kilday, Anne-Marie, and DS Nash. 2017. *Shame and modernity in Britain.* Springer.

Kimmel, Michael S. 2013. *Angry white men: American masculinity at the end of an era.* Nation Books.

Konow, James. 2000. “Fair shares: Accountability and cognitive dissonance in allocation decisions.” *American Economic Review* 90 (4): 1072–1091.

Kröhnert, Steffen, and Reiner Klingholz. 2006. *Not am Mann: von Helden der Arbeit zur neuen Unterschicht? Lebenlagen junger Erwachsener in wirtschaftlichen Abstiegsregionen der neuen Bundesländer.* 80. Berlin: Berlin-Institut für Bevölkerung und Entwicklung.

Le Yaouanq, Yves. 2021. *Motivated cognition in a model of voting.* Technical report. Working Paper.

Leeper, Thomas J, and Rune Slothuus. 2014. "Political parties, motivated reasoning, and public opinion formation." *Political Psychology* 35:129–156.

Leibert, Tim. 2016. "She leaves, he stays? Sex-selective migration in rural East Germany." *Journal of Rural Studies* 43:267–279.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2020. *A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data.* Technical report.

McCright, Aaron M., and Riley E. Dunlap. 2011. "The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010." *Sociological Quarterly* 52 (2): 155–194.

McDonald, Melissa M, Carlos D Navarrete, and Jim Sidanius. 2011. "Developing a theory of gendered prejudice." *Social cognition, social identity, and intergroup relations: A Festschrift in honor of Marilynn B. Brewer,* 189–220.

McDonald, Melissa M, Carlos David Navarrete, and Mark Van Vugt. 2012. "Evolution and the psychology of intergroup conflict: The male warrior hypothesis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1589): 670–679.

Medez, Ildefonso, and Isabell M. Cutillas. 2014. "Has immigration affected Spanish presidential elections results?" *Journal of Population Economics* 27:135–171.

Melzer, Silvia Maja. 2011. "Reconsidering the Effect of Education on East–West Migration in Germany." *European Sociological Review* 29, no. 2 (July): 210–228.

Messick, David M, and Keith P Sentis. 1979. "Fairness and preference." *Journal of Experimental Social Psychology* 15 (4): 418–434.

Miller, William Ian. 2002. *The mystery of courage.* Harvard University Press.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 0. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science* 0 (0): null.

Mohammad, Saif M., and Peter D. Turney. 2013. "Crowdsourcing a Word-Emotion Association Lexicon." *Computational Intelligence* 29 (3): 436–465.

Müller, Karsten, and Carlo Schwarz. 2020. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association* 19, no. 4 (October): 2131–2167.

Nickerson, Raymond S. 1998. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology* 2 (2): 175–220.

Nussbaum, Marta C. 2009. *Hiding from Humanity: Disgust, Shame, and the Law.* Princeton University Press.

Offer, Avner. 1995. "Going to War in 1914: A Matter of Honor?" *Politics & Society* 23 (2): 213–241.

Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics* 8 (1): 405–434.

Otterbein, Keith F. 1965. "Caribbean Family Organization: A Comparative Analysis." *American Anthropologist* 67 (1): 66–79.

Otto, Alkis Henri, and Max Friedrich Steinhardt. 2014. “Immigration and election outcomes — Evidence from city districts in Hamburg.” *Regional Science and Urban Economics* 45:67–79.

Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference.* Cambridge University Press.

Pedersen, Frank A. 1991. “Secular Trends in Human Sex Ratios - Their Influence on Individual and Family Behavior.” *Human Nature* 2:271–291.

Perez-Truglia, Ricardo, and Ugo Troiano. 2018. “Shaming tax delinquents.” *Journal of Public Economics* 167:120–137.

Rabin, Matthew, and Joel L Schrag. 1999. “First impressions matter: A model of confirmatory bias.” *Quarterly Journal of Economics* 114 (1): 37–82.

Robb, George. 2014. *British Culture and the First World War.* Macmillan International Higher Education.

Rusch, Hannes, Joost M. Leunissen, and Mark van Vugt. 2015. “Historical and experimental evidence of sexual selection for war heroism.” *Evolution and Human Behavior* 36 (5): 367–373.

Sampson, Robert J., John H. Laub, and Christopher Wimer. 2006. “Does marriage reduce crime? A counterfactual approach to within-individual causal effects.” *Criminology* 44 (3): 465–508.

Sapolsky, Robert M. 1990. “Adrenocortical function, social rank, and personality among wild baboons.” *Biological Psychiatry* 28 (10): 862–878.

Sapolsky, Robert M. 1991. “Testicular function, social rank and personality among wild baboons.” *Psychoneuroendocrinology* 16 (4): 281–293.

Schacht, Ryan, and Karen L. Kramer. 2016. “Patterns of Family Formation in Response to Sex Ratio Variation.” *PLOS ONE* 11, no. 8 (August): 1–14.

Schmitt, David P. 2005. “Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating.” *Behav Brain Sci.* 28 (2): 247–75.

Schwardmann, Peter, Egon Tripodi, and Joël J Van der Weele. 2021. “Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions.” *American Economic Review (forthcoming)*.

Schwardmann, Peter, and Joël J Van der Weele. 2019. “Deception and self-deception.” *Nature Human Behaviour* 3 (10): 1055–1061.

Shayo, Moses. 2020. “Social identity and economic policy.” *Annual Review of Economics* 12:355–389.

Solda, Alice, Changxia Ke, Lionel Page, and William Von Hippel. 2020. “Strategically delusional.” *Experimental Economics* 23 (3): 604–631.

Statistisches Bundesamt. 2018. *Bevölkerung und Erwerbstätigkeit, Schutzsuchende, Ergebnisse des Ausländerzentralregisters.* [Online; accessed 20-May-2022]. https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00041272/2010240167004_ueberarbeitet20122018.pdf.

Stauder, Johannes. 2018. "(Why) have women left East Germany more frequently than men?" *Heidelberger Jahrbücher Online* 3:73–97.

Steinmayr, Andreas. 2021. "Contact versus Exposure: Refugee Presence and Voting for the Far Right." *The Review of Economics and Statistics* 103, no. 2 (May): 310–327.

Stevens, Kimberly Elisa. 2016. "Flight of the white feather: The expansion of the white feather movement throughout the World War One British Commonwealth." *MA Thesis, U of Southern Georgia*.

Strezhnev, Anton. 2018. "Semiparametric weighting estimators for multi-period differencein-differences designs."

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 175–199.

Sznycer, Daniel, Dimitris Xygalatas, Elizabeth Agey, Sarah Alami, Xiao-Fen An, Kristina I. Ananyeva, Quentin D. Atkinson, et al. 2018. "Cross-cultural invariances in the architecture of shame." *Proceedings of the National Academy of Sciences* 115 (39): 9702–9707.

Taber, Charles S, and Milton Lodge. 2006. "Motivated skepticism in the evaluation of political beliefs." *American Journal of Political Science* 50 (3): 755–769.

Teso, Edoardo. 2018. "The Long-Term Effect of Demographic Shocks on the Evolution of Gender Roles: Evidence from the transatlantic Slave Trade." *Journal of the European Economic Association* 17, no. 2 (April): 497–534.

Thaler, Michael. 2020. "Do People Engage in Motivated Reasoning to Think the World Is a Good Place for Others?" *arXiv preprint arXiv:2012.01548*.

Van Bavel, Jay J, and Andrea Pereira. 2018. "The partisan brain: An identity-based model of political belief." *Trends in Cognitive Sciences* 22 (3): 213–224.

Voigtländer, Nico, and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany*." *The Quarterly Journal of Economics* 127, no. 3 (July): 1339–1392.

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

Von Bismarck-Osten, Clara, Kirill Borusyak, and Uta Schönberg. 2021. *The Role of Schools in Transmission of the SARS-CoV-2 Virus: Quasi-Experimental Evidence from Germany*. Technical report, CReAM Discussion Paper 22/20.

Vugt, Mark Van, David De Cremer, and Dirk P Janssen. 2007a. "Gender differences in cooperation and competition: The male-warrior hypothesis." *Psychological science* 18 (1): 19–23.

———. 2007b. "Gender Differences in Cooperation and Competition: The Male-Warrior Hypothesis." *Psychological Science* 18 (1): 19–23.

Ward, Paul. 2001. "'Women of Britain Say Go': Women's Patriotism in the First World War." *Twentieth Century British History* 12 (1): 23–45.

Wei, Shang-Jin, and Xiaobo Zhang. 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119 (3): 511–564.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1): 57–76.