COLLECTIVE ACTION AND SOCIAL CHANGE: HOW DO PROTESTS INFLUENCE SOCIAL MEDIA CONVERSATIONS ABOUT IMMIGRANTS?

by

Kai Wei

Bachelor of Arts, Northwestern Polytechnical University, 2011

Master of Social Work, University of Pittsburgh, 2014

Submitted to the Graduate Faculty of

School of Social Work in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SCHOOL OF SOCIAL WORK

This dissertation was presented

by

Kai Wei

It was defended on

May 23, 2019

and approved by

Jeffrey Shook, Ph. D, Associate Professor, School of Social Work

Yu-Ru Lin, Ph. D., Associate Professor, School of Computing and Information

Dissertation Co-advisor: Christina E. Newhill, Ph. D, Professor, School of Social Work

Dissertation Co-advisor: Jaime M. Booth, Ph. D, Assisant Professor, School of Social Work

COLLECTIVE ACTION AND SOCIAL CHANGE: HOW DO PROTESTS INFLUENCE SOCIAL MEDIA CONVERSATIONS ABOUT IMMIGRANTS?

Kai Wei, Ph.D.

University of Pittsburgh, 2019

Abstract

This study aims to improve our understanding of using protest as an intervention strategy to reduce online users' prejudiced speech against immigrants by: 1) developing a reliable method to measure for online users' prejudiced speech against immigrants; 2) examining the role of temporal and geographic exposure to protest in online users' prejudiced speech against immigrants; and 3) examining the role of group identity in the relationship between protest and online users' prejudiced speech against immigrants. It hypothesizes that protest would reduce online users' prejudiced speech against immigrants. After collecting 31,210,740 tweets from 102, 094 Twitter users, machine learning techniques were leveraged into developing a reliable measurement for online users' prejudiced speech against immigrants. Repeated measures of users' prejudiced speech were taken in a two-week window to establish baseline before and after protest. Analyses examined group differences across different levels of geographic exposure to local protest and between users identified with different groups using analysis of variance procedures. Overall, this research did not provide evidence supporting the claim that protest can reduce online prejudiced speech. However, it was found that users expressed more prejudiced speech after protest compared to baseline before protest. This change was more pronounced among users located furthest (in geographic distance) from the cities where protests occurred. It was also more pronounced among

users who did not identify with immigrants. Further research is needed to determine if these results call into question the effectiveness of protest in reducing prejudiced speech or are peculiar to social media, and if so, how these negative effects can be mitigated.

Table of Contents

List of Tables

List of Figures

DEDICATION

I dedicate this dissertation to

my beloved husband Kurt Wallnau,

who encouraged me to follow my dreams.

ACKNOWLEDGEMENTS

This work would not have been possible without the support of Dr. Christina E. Newhill and Dr. Jaime M. Booth, who inspired me to become an independent researcher. Thank you for being supportive of my research and career goals, and working actively with me to designate the academic time to pursue those goals. My sincere thanks must also go to Dr. Jeffrey Shook and Dr. Larry E. Davis. Thank you for believing in me before I believed in myself, and for providing me the opportunities to pursue social work research. I also would like to express my deepest appreciation to Dr. Yu-Ru Lin who opened the door to a completely new research world: leveraging computational methods to understand social problems. Thank you for providing me the invaluable training that made it possible for me to pursue cutting-edge research.

I would also like to extend my sincere thanks to my parents and other family members, many of whom have provided unfailing emotional support. I deeply thank my parents for their unconditional love and trust. It was their love that raised me up again and again when I wanted to give up. I also cannot thank my husband enough for his timely encouragement and unconditional love and support. Without you, I could not have made it this far. I also thank my father-in-law Carl Wallnau, and Liselott Mohr for being generous with your love and encouragement. I am also grateful to Varsha Ramasubramanian and Bryn Wallnau, whose inputs were indispensable in developing a reliable measurement in this dissertation. Last but not least, I would especially like to thank David Shelow, who provided vital comments that helped me improve my writing.

1.0 INTRODUCTION

Widespread expression of prejudice in public speech perpetuates, and in recent times has intensified, the feeling of immigrants of being excluded from full participation in society (Fangen, 2010; Mullen & Rice, 2003). As we have become an increasingly online society, prejudiced speech has also moved online, where it frequently takes on more overt and extreme forms (Davidson, Warmsley, Macy, & Weber, 2017). While previous research suggests that social protest can be effective in reducing stigmatizing public speech (Rüsch, Angermeyer, & Corrigan, 2005), research to date has not specifically examined the effects of social protest in online prejudiced speech. This dissertation seeks to determine the effectiveness of social protest in reducing online prejudiced speech against immigrants. This section begins with an introduction to the problem of prejudiced speech against immigrants, the role of social media in exacerbating this problem, and the need to examine the effectiveness of social protest as an intervention to reduce prejudiced online speech (§1.1). This is followed by a discussion of the significance of this study in social work research (§1.2). Finally, this chapter concludes with an overview of the specific studies carried out in this dissertation and research hypotheses with respect to each study (§1.3).

1.1 PROBLEM STATEMENT

The saying that "America is a nation of immigrants" has never been more true than it is today, when one in four Americans are immigrants or children of immigrants (Brown & Stepler, 2016). As a foreign-born population, immigrants face many challenges integrating into US society, including balancing conflicting cultural values and forming new social networks (Cervantes, Padilla, Napper, & Goldbach, 2013). Beyond these challenges, immigrants have historically confronted prejudice and discrimination conveyed in public speech, and this is still true for immigrants today (Wei, Jacobson López, & Wu, 2019). For example, on June 16, 2015, Donald J. Trump announced his presidential candidacy with a speech in which he explicitly referred to Mexican immigrants as "rapists" and "criminals."(Kohn, 2015). Previous research has shown that prejudice conveyed in this kind of public speech has led to immigrants' experiences of microaggression (Yosso, Smith, Ceja, & Solórzano, 2009), work-related discrimination (Hanna & Ortega, 2016), and hate crimes (Flores-Yeffal, Vidales, & Plemons, 2011). Moreover, it has intensified their feelings of being excluded from full participation in society (Fangen, 2010; Mullen & Rice, 2003).

While research has long been concerned with the corrosive effects of prejudiced speech, the advent of social media brings new urgency to this concern. Today, seven-in-ten Americans use social media to communicate with each other and share information (Pew Research Center, 2018). As our social activity takes place increasingly online, prejudiced speech has also moved online in an overt and extreme form (Davidson et al., 2017). For example, there has been growing concern about online users using social media as a tool to spread racist beliefs and to incite violence offline (Ben-David & Matamoros-Fernandez, 2016; Chan, Ghose, & Seamans, 2016). Although prejudiced speech occurs online, its adverse effects are demonstrably real offline, i.e., in the "real world" (Kowalski, Giumetti, Schroeder, & Lattanner, 2014). Online verbal attacks have been linked to unhealthy psychological developments such as depression, anxiety, and lower self-esteem (Kowalski et al., 2014). Immigrants are particularly susceptible to the real, adverse effects of online prejudice. As shown in a recent study, immigrant youths are more likely than US-born youths to become victims of online verbal aggression (Comas-Forgas, Sureda-Negre, & Calvo-Sastre, 2017).

As protest has been identified as a type of prejudice-reduction intervention (Corrigan & Penn, 1999), it might also be used as an intervention tool to reduce online prejudiced speech against immigrants. As suggested in previous research, protest can reduce prejudiced and stigmatizing public statements (Casados, 2017; Corrigan, Morris, Michaels, Rafacz, & Rüsch, 2012; Corrigan & Watson, 2002; Rüsch et al., 2005; Wahl, 1997). Exposure to protest has been found to produce positive attitudes toward racial minorities (Mazumder, 2018), more liberal gender attitudes (Banaszak & Ondercin, 2016), and awareness of immigration related issues (Carey Jr, Branton, & Martinez-Ebers, 2014). For example, Americans from counties that experienced civil rights protests were found to express more positive attitudes towards racial minorities compared with those from counties that did not experience protests (Mazumder, 2018). While previous research has shown that social protests can reduce prejudiced speech and improve attitudes, research to date has not specifically examined the effects of social protest on online users' prejudiced speech.

1.2 RELEVANCE TO SOCIAL WORK

Advocacy is a core activity of social work practice, distinguishing social workers from other helping professions, and originating in social work's emphasis on understanding people within their environment (Kirst-Ashman & Hull, 2002). In social work with immigrants, Balgopal (2000) suggested that "the role of the social worker is to learn how to assess immigrants' situations, advocate for their rights and needs, determine which community resources they need, help them adapt to their new country without leaving behind their cultural customs and traditions, and monitor their progress" (p. 238). This requires social work with immigrants not only to help immigrants integrate into US society, and deliver social services, but also to advocate for immigrants' rights and needs.

Advocacy on behalf of immigrants and empowering them to advocate for themselves is especially important for today's social work with immigrants. This is because the contemporary anti-immigrant political climate not only intensifies the social, political, and economic inequalities facing immigrants, but also creates ethical dilemmas for social workers who work closely with immigrants. Today, most social workers are employed by state or local governments, and their delivery of services to immigrants is regulated by immigration policy (Furman, Langer, Sanchez, & Negi, 2007). For example, Proposition 200 in Arizona requires social workers to check the legal status of individuals prior to providing any public benefits. In this case, social workers are caught between the law, which denies services to undocumented clients, and the Code of Ethics, which mandates the provision of services to all vulnerable populations (Furman et al., 2007). To resolve ethical dilemmas, practical social work with immigrants must emphasize advocacy for social change.

Social workers have used various forms of advocacy such as lobbying, letter-writing, and protest to address social injustice. For example, the National Association of Social Workers recently published a statement opposing President Donald Trump's executive order that banned Muslim immigrants from entering the United States (NASW, 2017). In Tacoma, social workers designed a project to raise awareness of immigrants' experiences in the United States by gathering stories from detention officers who have witnessed how families have been torn apart under the enforcement of current immigration policies (Epps, 2015). While protest is an important and effective advocacy tool in social work, it remains an under-researched field in social work (Moshe Grodofsky & Makaros, 2016). Therefore, more research is needed to advance our knowledge of using protest as an advocacy tool.

Recent research suggests that social workers can harness social media for scaling up advocacy works and bringing meaningful social change (Brady, Young, & McLeod, 2015; Gandy-Guedes, Vance, Bridgewater, Montgomery, & Taylor, 2016; Greeson, An, Xue, Thompson, & Guo, 2018; Hitchcock & Young, 2016; López Peláez, Pérez García, & Aguilar-Tablada Massó, 2018; Sitter & Curnew, 2016). Sitter and Curnew (2016) suggested that social workers can use social media to promote social justice and advocacy along with community groups, amplify voices for marginalized groups, mobilized supports, and foster community connections online and offline. In addition to these functions of social media, social work can also exploit social media to promote and organize collective action. For example, Brady et al. (2015) described an organizing effort undertaken by social work academics and allies that involved online organizing to support low-wage workers for fair pay and better working conditions. There is also evidence that using social media to share personal experiences can transform public perceptions of marginalized groups such as the lesbian, gay, bisexual, and transgender community (Jones, 2015). These above examples highlight the potential of using social media as a tool for advocacy and addressing issues confronted with immigrants.

However, naive use of social media to support advocacy can pose significant risks. For example, social media platforms give rise to “filter bubbles”, where people are given information that is already aligned with their views of the world and “echo chambers,” where people share information only with like-minded people (Bakshy, Messing, & Adamic, 2015; Pariser, 2011). The combination of user anonymity, attention seeking habits, and other poorly understood dynamics of online social behavior can result in inflamed passions, and inadvertently lead to more intense prejudiced speech against immigrants. Shared personal experiences are more likely to be seen only by, and shared only with, social media users whose perceptions are already aligned with these worldviews (Boutyline & Willer, 2017). Penetrating filter bubbles and echo chambers may require more disruptive forms of advocacy, such as social protest. However, disruptive forms of online advocacy are fraught with legal and ethical concerns. As one example, de-anonymizing social media users in the service of public shaming constitutes a violation of terms of use license agreements and possibly statute (Wondracek, Holz, Kirda, & Kruegel, 2010). Therefore, more research is needed to help social workers harness social media to bring about meaningful social change.

1.3 OVERVIEW OF THIS DISSERTATION

The goal of this dissertation is to improve our understanding of using protest as an intervention strategy to reduce online prejudiced speech against immigrants. As discussed previously, persistent problems with prejudice conveyed in public speech have perpetuated and intensified immigrants' experiences of being excluded from full participation in society (Fangen, 2010; Mullen & Rice, 2003). As we are increasingly living in online society, prejudiced speech has also moved online in an overt and extreme form (Davidson et al., 2017). Today, one in seven Americans use social media to connect with each other and share information, and many also use social media to promote protest and collective action (Pew Research Center, 2017). While prejudice-reduction research has suggested the importance of social protest in reducing stigmatizing public statements (Rüsch et al., 2005), research to date has not examined the role of protest in online prejudiced speech. Therefore, this dissertation carried out three studies to examine the role of protest in online prejudiced speech.

1.3.1 Study aims and research hypotheses

This study aims to improve our understanding of using protest as an intervention strategy to reduce online users' prejudiced speech against immigrants by:

1) developing a reliable method to measure for online users' prejudiced speech against immigrants (**study aim #1)**;
2) examining the role of temporal and geographic exposure to protest in online users' prejudiced speech against immigrants (**study aim #2)**;
3) examining the role of group identity in the relationship between protest and online users' prejudiced speech against immigrants (**study aim #3)**.

To achieve **study aim #1,** I developed a method for measuring online users' prejudiced speech against immigrants. One of the challenges in studying the role of protest in online prejudiced speech is to obtain reliable outcome measurements. While prior work proposed methods for measuring online hate speech, these methods have ignored the targets of hate speech (ElSherief, Kulkarni, Nguyen, Wang, & Belding, 2018) and are consequently not applicable for measuring online prejudiced speech against immigrants in particular. To address this challenge, this study leveraged machine learning techniques to automatically label tweets. The key idea of this method is to identify a collection of tweets that are relevant to immigrants, from which statistical algorithms are used to automatically classify whether a tweet is about immigrants and whether a tweet is prejudiced speech against immigrants. Users' prejudiced speech is then measured based on the proportion of tweets that are classified as prejudiced speech against immigrants to tweets that are classified as about immigrants (see §4.4.3 for details about methods description and §5.1 for measure accuracy results).

To achieve **study aim #2,** I examined the role of temporal and spatial exposure to protest in online prejudiced speech against immigrants. Protest has been identified as one of the major prejudice-reduction interventions (Casados, 2017; Corrigan, Morris, Michaels, Rafacz, & Rüsch, 2012; Corrigan & Watson, 2002; Rüsch et al., 2005; Wahl, 1997; detailed discussions of protest is included in §2.3). Social movement impact theory poses that protests can lead to changes in political and policy outcomes (Gamson, 1975; Piven & Cloward, 1977) and can influence cultural outcomes (Bernstein, 2003; see details in §3.1). Previous studies focusing on the relationship between protest exposure and attitude changes have shown that exposure to protest can lead to improvement in gender attitudes, and reduction in racial prejudice and negative attitudes against people with mental illness (Banaszak & Ondercin, 2016; Boyle, Dioguardi, & Pate, 2016;

Mazumder, 2018; see detailed discussions of protest exposure and attitude change in §2.3.2). Using the case of the American civil right movement, Mazumder (2018) developed a theory of civil resistance and a long-run attitude change which poses that instances of collective action by a protest movement can generate attitude change among the target public (see §3.2 for detailed discussion). Together, these studies suggest that protest may reduce online prejudiced speech against immigrants. Therefore, it was hypothesized that:

> <u>Hypothesis 1:</u> Social media users express less prejudiced speech against immigrants after protests compared to baseline before protests.
>
> <u>Hypothesis 2:</u> Higher levels of exposure to local protest are associated with less prejudiced speech against immigrants among social media users.

To achieve **study aim #2,** I examined the role of group identity in the relationship between protest and online prejudiced speech against immigrants. Social identity informs people about belonging or not belonging to a specific group (Goldenberg, Halperin, van Zomeren, & Gross, 2016). It transforms people's self-perception, feelings, and behavior to conform to a group normative response to an out-group (Turner & Reynolds, 2011), and it is likely to be activated in situations of intergroup conflicts, rivalry, or social comparison (Smith & Mackie, 2015). Past research on prejudice has linked people's social identity to prejudice against immigrants and racial minorities (Dru, 2007; Oswald, 2005; Ray, Mackie, Rydell, & Smith, 2008). The overall findings show that participants would express more prejudice when they did not identify with others (see §2.2.2 for detailed discussions of social identity and prejudice). According to intergroup emotion theory, when an event negatively influences a group, people identified with that group will interpret the event negatively and have negative emotional reactions toward the event, even if they are not directly harmed (see §3.3 for detailed discussions of intergroup emotion theory). Together, these studies suggest that compared to out-group members, in-group members might express less prejudiced speech after protest. Therefore, it was hypothesized that:

> Hypothesis 3: After the protest, social media users who were self-identified as immigrants express less prejudiced speech against immigrants compared to users who were self-identified as non-immigrants.

1.3.2 Major contributions of this dissertation

The results of this dissertation provide new insights into the role of social protest in social media prejudiced speech against immigrants. Specifically, this dissertation presents a reliable method for measuring online users' prejudiced speech, provides empirical evidence for the relationship between users' geo-locations to protest cities and their online prejudice speech after protest events, improves the theoretical understanding on the role of protest and group identity in online prejudice speech, and demonstrates the utility of computational focus group framework for understanding online and offline dynamics. In the long run, this dissertation will help social workers understand the advantages and disadvantages of using protest as a prejudice-reduction intervention and providing insights for making better use of online platforms to amplify the intended effects of protest.

1.3.3 Terminology

I use the term *immigrant* in this dissertation to refer to persons who have migrated to the US on a long-term basis, which includes those who are, or are in the process of becoming, naturalized citizens or lawful permanent residents; those who are refugees or asylum-seekers; and those who reside in the country without government permission (Zong & Batalova, 2015).

This dissertation adopts Allport's (1954) definition of prejudice as individuals' antipathy towards a person or a group. It is based on a feeling of dislike, false assumptions, and stereotypes about a personal or a group. This dissertation focuses on prejudice conveyed in public speech, i.e., social media platforms. Prejudiced speech is defined as the expressions of hostility in one's public statements about a person or a group. For example, a person hold prejudice against immigrant can make the following statement about immigrants to express their dislike about immigrants in his or her social media post, "#DayWithoutImmigrants is great, DECADES without immigrants would be even better, they should all go back to their home".

2.0 LITERATURE REVIEW

Prejudice is a feeling of dislike based on false assumptions and stereotypes, and can be expressed in speech. Prejudiced speech can directly reflect people's prejudices and lead to discrimination, social exclusion, and even physical assault. In the past several decades, there has been a marked increase of research activities devoted to understanding the causes of prejudice and ways to reduce it. This chapter provides a review of the literature to highlight the central issue of prejudiced speech against immigrants, beginning with an overview of prejudiced speech and social exclusion of immigrants in the United States (§2.1). This is followed by an examination of preliminary evidence that identifies social identity as a contributor to prejudiced speech against immigrants (§2.2). Third, it proceeds with a detailed review of prejudice-reduction interventions, with an eye toward the role of protest in prejudice reduction (§2.3). Fourth, it reviews the emergence of social media and its role in prejudice (§2.4). Finally, this chapter concludes with an overview of current studies, highlighting the importance of understanding the relationship between protest and online prejudiced speech against immigrants, and the role of social identity in moderating the relationship between the two.

2.1 PREJUDICE AND SOCIAL EXCLUSION OF IMMIGRANTS

People often convey their prejudices through private or public speech (Allport, 1954), and when they do so, it does not merely express the negative feelings of the speaker; speech acts on listeners to separate, and create distance, between the listeners and the group targeted by prejudiced speech

(Austin, 1975; Graumann & Wintermantel, 1989). Previous studies have linked prejudice conveyed in public speech to immigrants' experiences of microaggression (Yosso et al., 2009), work-related discrimination (Hanna & Ortega, 2016), hate crimes (Flores-Yeffal et al., 2011), and social exclusion (Mullen & Rice, 2003). These negative consequences of prejudiced speech are especially important for understanding social exclusion of Muslims and Mexicans. Both groups have been subjected to simplified and negative portrayals in public discourse about immigration, with all Muslims portrayed as terrorists, and all Mexicans portrayed as illegals.

Mexican immigrants (regardless of their legal status) and Mexican Americans often encounter verbal insults from people who assume they are illegal (Firzsimon, 2007; Short & Magaña, 2002). At work and in everyday life, Mexican immigrants are not only underpaid for the same work, but also are openly ridiculed by people who assume they do not understand English well enough to appreciate the insults being hurled at them (Hanna & Ortega, 2016). Muslim immigrants and Muslim Americans can experience similar forms of marginalization, causing a growing number of scholars to express concern about the rise of *Islamophobia* (hatred towards anyone of Islamic belief) toward Arabs and Muslims (Mastnak, 2008; Rana, 2007; Sheehi, 2011). In the past two decades, Arabs and Muslims have been the targets of an increasing number of hate crimes (Alsultany, 2012, 2013; FBI, 2001) and bias-motivated incidents such as airline security profiling, physical assaults of students wearing the hijab, arson and vandalism of mosques, and death threats and shootings at Muslim-owned business (Lichtablau, 2015).

While Mexicans and Muslims have been the primary targets of prejudice towards immigrants in recent times, their experiences are not unique in US history. Using archival data gathered over the course of 150 years, Mullen and Rice (2003) showed that immigrant groups regardless of nationality or ethnicity were depicted in simplified ways and subjected to negative ethnic slurs; they were also more likely to be segregated into ethnic neighborhoods, deemed suitable only for hazardous work or menial labor, and subjected to harsher immigration quotas. Moreover, in periods when these forms of prejudice were more pronounced, immigrants were less likely to become naturalized citizens. These findings from historical data are also supported by more recent experimental research (Leader, Mullen, & Rice, 2009) in which Leader and colleagues (2009) studied how exclusion of ethnic out-groups is related to prejudiced speech. They found that the simpler the sentences participants used to express prejudice against an ethnic out-group, the more likely they were to want to exclude the out-group from society, such as not allowing the out-group to have citizenship in their country.

Prior to 1965, decisions about which immigrant groups to include or exclude depended largely on their racial categories. Race-based laws were commonplace and were sustained by *scientific racism* (Dred Scott v. Sandford, 1857). From the 1820s to 1890s, while most Irish immigrants were naturalized and became Americans citizens, most Chinese immigrants were excluded from citizenship and barred from entering the country (Pub. L. No. 47-126, 1882). Chinese immigrants were perceived as racially inferior and as a threat to the purity of the white American identity. Racial prejudice against Chinese immigrants was conveyed in prejudiced speech and legal rulings with terms such as *Asiatic blood* and *yellow peril* (Young, 2014). For example, the lynching of 18 Chinese immigrants on October 24, 1871, in Los Angeles, California was never punished because a California Supreme Court judge ruled that Chinese witnesses could not testify against white people because Chinese were an inferior race with insufficient intelligence (People v. Hall, 1854). Race-based social exclusion of immigrants from American society also existed for other racial minorities. During the 1930s, Mexican repatriation removed more than 400,000 Mexican American citizens (California Senate Bill No. 670., 2005). During World War

II, while Germany and Japan were both enemies of the United States, only people of Japanese ancestry were relocated to internment camps (Ng, 2002).

In addition to exclusion from citizenship, racial minority immigrants were also excluded from owning or leasing agricultural land, and the exclusions ultimately expanded to include all real property such as sharecropping contracts and shares of stock in corporations owning agricultural land (Ferguson, 1947). These laws were driven by racial prejudice (Higham, 2002) and relegated racial minority immigrants to a status of perpetual aliens, foreigners, and sojourners (Young, 2014). Before the 1950s, Japanese and Chinese immigrants were deemed as unfit to be Americans because of their racial identity, and were systematically excluded from citizenship and subjected to occupational discrimination and residential segregation (Wu, 2014).

Although the 1965 US Immigration and Nationality Act eliminated explicit racial bias in immigration policy (The US Immigration and Nationality Act, 1965), race and ethnicity still remain central to the process of determining who is included, and who is excluded, from becoming American (Viruell-Fuentes, Miranda, & Abdulrahim, 2012). The shift from explicit (overt) to implicit (covert) racial bias in immigration is reflected in the shift from derogatory terms for Mexican immigrants such as "wetback" that were widely used in major newspapers during 1950s and 1960s, to covertly prejudiced terms such as "illegal alien" (Ackerman, 2013). This is not to say that no progress has been made in eliminating racial and ethnic bias in language about immigrants. By the 1990s, the term "illegal alien" was widely seen as covertly prejudiced (Florido, 2015). Today, the terms "illegal immigrant" and more recently "undocumented immigrant" has become a more widely accepted term in public discourse, though both terms remain derogatory.

The shift from overt to covert expressions of racial prejudice merely obscures the taint of racial bias while still serving to justify unequal treatment, and allow policies and social practices to exclude a particular racial or ethnic group from American society. For example, the enactment of the Personal Responsibility and Work Opportunity Act of 1996 (PRWORA) and the Affordable Care Act of 2010 (the ACA) restricted access to public assistance and health care on the basis of immigration status (Public Law 111-148, 2010; The Personal Responsibility and Work Opportunity Reconciliation Act, 1996). The enactment of the Personal Responsibility and Work Opportunity Act of 1996 (PRWORA) was a turning point, making immigration status a prerequisite for health entitlements among immigrants. It was the first federal law that restricted access to health care on the basis of immigration status by prohibiting access to Medicaid and the Supplemental Nutrition Assistance Program (SNAP) for undocumented immigrants, and also excluded legal permanent residents with fewer than five years residency (The Personal Responsibility and Work Opportunity Reconciliation Act, 1996).

Social welfare exclusion was extended in the Affordable Care Act (the ACA), which was specifically targeted to benefit low-income individuals and families. The stigmatizing narrative of undocumented immigrants abusing American welfare was explicit in debates over the health care bill on the floors of the US Congress. As a result, low-income immigrants were disproportionately excluded from this health care effort. In 2011, approximately 10 million legal immigrants and 5.5 million undocumented immigrants lacked healthcare insurance (Arredondo, Orozco, Wallace, & Rodríguez, 2012; Kaiser Family Foundation, 2013). Almost all immigrants excluded from the ACA benefits on the basis of their immigration status would have otherwise qualified on the basis of their low income (Kaiser Family Foundation, 2013).

The ultimate form of social exclusion of undesirable immigrants is deportation[1]. From 2004 to 2013, the deportation of unauthorized immigrants with no criminal background increased from

1. Deportation was first used to exclude Chinese immigrants (1891 Immigration Act, 1891).

148,285 to 240,027 (USDHS, 2014). During the same period, the deportation of unauthorized Mexican immigrants with no criminal background accounted for more than 70% of all deported unauthorized immigrants (USDHS, 2014). Deportation permitted massive raids without warrants in the search for unauthorized immigrants, and created an intimidating situation that disproportionately affected Mexican Americans and Central American immigrants (Aldana, 2007; Romero, 2006). This "conveys a message of exclusion, of being unwelcome" (López, 2012, p. 159) not only to unauthorized Mexican immigrants, but also to first and second generation Mexican Americans. For example, President Trump declared Mexican American Judge Gonzalo Curiel unfit to make a legal ruling because he was born to Mexican immigrants (Graham, 2016).

The social exclusion of immigrants (both legal and unauthorized) becomes more apparent after the terrorist attacks of 9/11. The negative portrayals of immigrants being criminals extended to them also being terrorists (Romero, 2008). One notable change is that the US Citizenship and Immigration Services (USCIS), which existed before 9/11, became part of Department of Homeland Security after 9/11, suggesting that the regulation of citizenship has officially become a national security issue. Following 9/11, Congress passed a series of immigration laws, ostensibly for national security purposes, among which the USA Patriot Act of 2001 (USAPA) might be the most controversial. The USAPA has been criticized for its potential violation of individuals' constitutional rights, such as freedom of religion and due process of the law, and for its devastating impact on the Muslim community (Whitehead & Aden, 2001; Wong, 2006). In 2003, conflation of "unauthorized immigrant" with "terrorist" is again seen in Operation Endgame, conducted by Immigration and Customs Enforcement (ICE), to detain and deport all unauthorized immigrants and suspected terrorists living in the United States (DHS, 2003). The Border Protection, Anti-terrorism, and Illegal Immigration Control Act of 2005, increased penalties for unauthorized immigrants by classifying unauthorized immigrants and anyone who assisted them in entering or remaining in the US as felons (H.R. 4437, 2005). Most recently, the presidential executive orders banning Muslim populations from entering the United States and building a wall on the US-Mexico border reinforce prejudice against Muslim and Mexican immigrants, and provide justification for excluding them from participating in American society.

2.2 PSYCHOLOGICAL AND SOCIAL FACTORS IN PREJUDICE

A recent increase in research activities has been devoted to understanding the psychological and social factors in prejudice, focusing on three basic areas: authoritarianism, social norms, and social identity. This section reviews the literature from these three areas in order to highlight the need to further study the role of social identity in prejudice against immigrants. Given the limited research investigating prejudice against immigrants in the United States (Zárate & Quezada, 2012), this review also includes prejudice studies conducted in settings outside of the United States.

2.2.1 Authoritarianism and social dominance orientations

In the aftermath of World War II, many asked: how could such a horrendous thing as the holocaust have occurred? Many psychologists sought to explain antisemitism as resulting from authoritarianism and social dominance orientations. Authoritarianism is an attitude or belief promoting absolute obedience or submission to authority (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950; Fromm, 1941; Maslow, 1943; McClosky, 1958; Siegel, 1956). Social dominance orientation is closely related to authoritarianism but remains a distinct construct (Duckitt & Sibley, 2007; Sidanius & Pratto, 2001) emphasizing the role of individual preference for group-based hierarchy and in-group dominance (Pratto, Sidanius, Stallworth, & Malle, 1994). This perspective postulates that societies minimize group conflicts by creating consensus on ideologies that promote the superiority of one group over others (Sidanius, Pratto, Martin, & Stallworth, 1991).

Particularly at this time in the United States, it is important to understand the role of authoritarianism and social dominance orientation in prejudice against immigrants. Recent research shows that Trump's campaign speeches used "us-versus-them" rhetoric, which appealed greatly to authoritarian Americans (MacWilliams, 2016). In Trump's speeches, the absolutist terms such as "losers" and "complete disasters" were classic authoritarian rhetoric, and his distinction between groups at the top of society (White) and those at the bottom (immigrants, Blacks, and Latinos) was a classic social dominant stance (Pettigrew, 2017). Trump's speeches broke the norms and conventions that suppressed prejudiced speech, and made it acceptable to express prejudices openly and freely.

Authoritarian and social dominant rhetoric serves to justify people's prejudices and promote prejudiced speech not only against immigrants, but other minority groups as well, as shown in a recent study that "many out-group prejudices characterize dedicated Trump followers, not just anti-immigrants, but anti-out-groups in general" (Pettigrew, 2017, p. 109). In fact, people who express prejudice against immigrants also tend to do so toward women, blacks, sexual minority groups, and people with disabilities (Akrami, Ekehammar, & Bergh, 2011; Ekehammar & Akrami, 2003). This phenomenon has been documented as generalized prejudice (McFarland, 2010), where authoritarianism and social dominance orientation have been found to be related to generalized prejudice.

People with authoritarian and social dominance traits tend to organize their worldviews based on their social identity, and perceive other groups as threats to their values and as morally deficient (Altemeyer, 1981 1988). Previous studies have provided consistent evidence that people who rank high in authoritarianism and social dominance orientation tend to express prejudice against ethnic minorities and immigrants (Asbrock, Christ, Duckitt, & Sibley, 2012; Cohrs & Stelzl, 2010; Echebarria‐Echabe & Guede, 2007; Hiel & Mervielde, 2005; Hodson & Esses, 2005; Imhoff & Recker, 2012; Van Assche, Roets, Dhont, & Van Hiel, 2014; Whitley Jr, 1999; Zakrisson, 2005). For example, Kugler, Jost, and Noorbaloochi (2014) found that American participants who endorsed authoritarianism were more likely to believe that the lives of Muslims immigrants are not as valuable as the lives of Americans, that immigrants do not make good Americans, and that American-born children of immigrants should not be given citizenship. These findings seem to suggest that authoritarianism and social dominance orientation alone might not be enough to explain people's prejudices against immigrants, and that social identity may play a more fundamental role in understanding prejudice against immigrants.

2.2.2 Social identity

People maintain many identities in their everyday lives (Lazarus, 1982), and these collective identities define what it means to be who they are (Burke & Tully, 1977), not only based on their personal attributes, but also depends on the groups with which they identify (Tajfel & Turner, 1979). Social identity informs people about belonging or not belonging to a specific group (Goldenberg, Halperin, van Zomeren, & Gross, 2016). A group can be a small number of people who interact face-to-face, such as a football team or working group, or a large number of people who share common characteristics such as nationality, race/ethnicity, and political ideology. When any of these groups becomes salient in their sense of group belonging, people think of themselves as part of in-group members ("we") and others as out-group members ("they"), rather than as unique individuals. Social identity transforms people's self-perception, feelings, and behavior to conform to a group normative response to an out-group (Turner & Reynolds, 2011), and it is likely to be activated in situations of intergroup conflicts, rivalry, or social comparison (Smith & Mackie, 2015).

Previous studies have linked people's social identity to prejudice toward Arab, Asian, Turkish, and black people (Dru, 2007; Oswald, 2005; Ray, Mackie, Rydell, & Smith, 2008). The overall findings suggest that participants would express more prejudice when their national identity was made salient. For example, Ray and colleagues (2008) showed that American students who self-categorized themselves as Americans expressed more antipathy towards Muslims compared to those who self-categorized themselves as students. Moreover, there is evidence that social identity has more influence on people's prejudice than their authoritarianism and social dominance orientation. A study conducted in France found that the relationship between prejudice toward immigrants and authoritarianism was conditioned on the salience of participants' national identity (Dru, 2007). Similarly, the impact of authoritarianism on prejudice was insignificant when Dutch and Australian students were given information that primed national identity (Verkuyten & Hagendoorn, 1998).

Reynolds and colleagues (2001) further showed that the effect of authoritarianism on prejudice was reduced when participants' identity reflected high consensus and common position with anoutgroup. This postulate was supported by a later study conducted in Canada, where Esses, Wagner, Wolf, Preiser, and Wilbur (2006) found that Canadian participants showed more positive attitudes towards immigrants when they were presented with information that primed a national identity that included immigrants, compared to participants who were presented with information that heightened the salience of national identity that excluded immigrants. More recently, research in Switzerland found that participants were more likely to express prejudice against immigrants who were perceived as a homogeneous group (Falomir-Pichastor & Frederic, 2013). They further suggested that promoting a heterogeneous in-group identity and an image of immigrants as heterogeneous while simultaneously encouraging an environment that could be more inclusive of immigrants who wish to integrate the in-group (Roblain, Malki, Azzi, & Licata, 2017) may help reduce prejudice and discrimination against immigrants.

These previous studies on the connection between social identity and prejudice provide insights on prejudice against immigrants among minority Americans. Bobo and Hutchings (1996) found that compared to white American participants, African American participants reported higher levels of perceived economic threats from Asian and Latino immigrants. Research on *black nativism* supported this finding, showing African Americans hold prejudice toward immigrants regardless of their racial and ethnic identities (Breitzer, 2011). In addition to African Americans, Latino Americans and Mexican Americans were also found to express prejudice toward Mexican and Muslim immigrants. Hitlan, Carrillo, Zárate, and Aikman (2007) found that following the September 11, 2001 terrorist attack, Latino participants were more likely to perceive Muslim immigrants as a cultural threat and to perceive Mexican immigrants as an economic threat than they were before the terrorist attack. Among Mexican Americans, low-income Mexican Americans perceived undocumented Mexican immigrants as a threat to their job security; and high-income Mexican Americans perceived Mexican immigrants as a threat to their access to public education and the welfare system (Jiménez, 2007; Miller, Polinard, & Wrinkle, 1984; Schwartz, 1999). These findings showed that in situations of conflict, racial and ethnic minorities may express prejudice toward immigrants who share the same racial and ethnic identities because these situations primed their group attachment to American identity, rather than their racial and ethnic identities.

Given that immigration plays a large part in contemporary US political debate, political identity also has been linked to prejudice toward immigrants. People who self-identified as conservatives were more likely to express prejudice toward immigrants compared to those who self-identified as liberals (Caricati, Mancini, & Marletta, 2017; Chambers, Schlenker, & Collisson, 2013; Kugler et al., 2014; Van de Vyver, Houston, Abrams, & Vasiljevic, 2016). Previous researchers also found that affirming a conservative identity informs a belief system that emphasizes social dominance, which in turn can justify people's prejudices toward immigrants as well as minority Americans (Bobo, 1997; Kinder & Sears, 1981; Kluegel & Smith, 1986; Sears, 1988; Sidanius, Pratto, & Bobo, 1996). The link between conservative identity and social dominance orientation has been found in conservative political rhetoric; for example, conservative rhetoric is often accompanied by an emphasis on dangerous and competitive worldviews (Lakoff, 1997; Lane, 1962).

While previous studies have primarily suggested the positive relationship between conservative identity and prejudice toward immigrants, recent research found that people who self-identified as liberals can also display prejudice toward immigrants (Jost, Stern, Rule, & Sterling, 2017; Van de Vyver et al., 2016). There is evidence that self-identified liberals reported higher levels of prejudice toward Muslims and immigrants following a terrorist attack. This may be due to the increase in their concerns about in-group loyalty and authority (Van de Vyver et al., 2016), or their psychological reaction to fear and threat (Jost et al., 2017).

2.2.3 Social norms

Social identity is informed by group norms and can transform individuals' feelings and behavior to conform to a group normative response, such that members of a group develop prejudice norms through interacting with each other, and they pressure each other to conform to these norms (Sherif & Sherif, 1953). Prejudices that people express against immigrants may simply be a way for members of a group to indicate to their fellow members that they think it is socially acceptable. Therefore, it is important to review how social norms plays a role in prejudice against immigrants.

Social norms inform people about what is acceptable and unacceptable group behavior (Cialdini, Kallgren, & Reno, 1991; Cialdini, Reno, & Kallgren, 1990; Shaffer, 1983). Early studies have shown that social norms play an important role in shaping prejudice (Bolton, 1935; DeFriese & Ford, 1969; Hamblin, 1962; Lewin, 1947; Lippitt, 1949; Marrow & French, 1945; Pettigrew, 1958, 1959, 1960). When negative stereotypes are socially agreed, people often speak in prejudiced ways (Katz & Braly, 1933), because they learn and conform to prejudiced norms (Sherif & Sherif, 1953) through observation and communication (Sechrist & Stangor, 2004). This process may explain why people feel different emotions toward immigrants when they self-identify as one group versus the other.

People closely adhere to social norms when expressing prejudice and reacting to hostile jokes (Crandall, Eshleman, & O'brien, 2002). Hearing one person condemn or condone racial prejudice can produce more condemning or condoning reactions regardless of the person's social identity (Blanchard, Crandall, Brigham, & Vaughn, 1994). A series of early studies on racial prejudice have shown that participants who self-identified as European Americans expressed much less racial prejudice to a Black interviewer than they did to a White interviewer (Kinder & Sanders, 1996; Lowery, Hardin, & Sinclair, 2001; Schuman, 1997). Consistent with this finding, recent research showed similar results in the study of prejudice toward immigrants. In an experimental study conducted in Norway, researchers found that participants expressed fewer prejudiced statements about immigrants in group discussions with a non-Norwegian interviewer than they did with a Norwegian interviewer (Klöckner & de Raaf, 2013). These studies support the role of social norms in prejudice toward immigrants, and suggest that the way people express prejudice depends on their perception of whether or not that behavior is acceptable in a given context.

Research examining social norms and prejudice also suggests that the clarity of social norms also plays an important role in how people express prejudice (Zitek & Hebl, 2007). In today's United States, while the prevalent social norm is that open expression of racial and ethnic prejudice is unacceptable (França & Monteiro, 2013; Klöckner & de Raaf, 2013), several scholars have argued that prejudice toward immigrants is or has become socially acceptable (Fisher, Deason, Borgida, & Oyamot, 2011; Nier, Gaertner, Nier, & Dovidio, 2012; Zárate & Quezada, 2012). This line of research also linked socially-sanctioned prejudice toward immigrants to current immigration laws, suggesting that legal enforcement against immigration might be Americans' way of expressing prejudice toward immigrants (Zárate & Quezada, 2012).

Given that the expression of prejudice is sanctioned both by norms and laws (Fisher et al., 2011), it is not surprising that most immigration laws aim to punish unauthorized immigrants, but enact no consequences for Americans who are engaged in the unlawful employment of unauthorized immigrants (Mukherjee, Molina, & Adams, 2012). Recent studies have provided support for this treatment deferential to Americans (Hartley & Armendariz, 2011; Light, Massoglia, & King, 2014; Wolfe, Pyrooz, & Spohn, 2011). Analyzing nearly two decades of federal court records to determine whether citizenship had implications for different legal treatment of American citizens and both legal and unauthorized immigrants, Light et al. (2014) found that citizenship status plays a more powerful role in predicting sentencing outcomes than the race or ethnicity of immigrants. Compared to American citizens, legal immigrants are twice as likely to be imprisoned, and unauthorized immigrants are seven times more likely to be imprisoned. On average, legal immigrants received an additional 3.5 months of prison time than American citizens for the same offense, after controlling for criminal history, race and ethnicity. The incarceration gap between American citizens and immigrants is wider in districts with higher percentages of immigrant populations. Moreover, the effect of citizenship on sentencing has become more pronounced in recent years: in 1992, immigrants were twice as likely than American citizens to be incarcerated by a federal judge; by 2008, they were four times as likely than American citizens to be incarcerated.

2.3 PROTEST AND PREJUDICE-REDUCTION

Over the past several decades, psychology literature has identified various interventions that reduce prejudice and stigma. These interventions, according to Corrigan and Penn (1999), can be categorized as contact (facilitating positive inter-group contact), education (replacing myths about an out-group with accurate knowledge), and protest (attempts to suppress prejudice against an out-group). The following section reviews the current state of research in prejudice-reduction, and discusses the need for examining the role of protest in reducing prejudice against immigrants.

2.3.1 Contact and education

The most widely used approach in prejudice-reduction is contact, which allows members of different groups to resolve conflicts by communicating and learning to appreciate different points of view. This approach is based on inter-group contact theory (Allport, 1954), the premise of which is that people are prejudiced because they feel anxious about unfamiliar out-group members (Stephan & Stephan, 2000). Therefore, promoting intergroup interaction would help people overcome their anxiety, and thereby reduce their prejudices. Given that prejudice toward ethnic minorities begins during the ages of three to six years (Raabe & Beelmann, 2011), studies testing the effect of contact have primarily focused on children (Aboud & Miller, 2007; Feddes, Noack, & Rutland, 2009; Gaertner, Mann, Murrell, & Dovidio, 1989). Previous studies have shown mixed effects of contact-based interventions on reducing prejudice. There was evidence that after promoting positive contact, participants overall increased their positive perceptions about out-group members (Pettigrew & Tropp, 2006), placed more emphasis on personal identities rather than social identities (Brewer & Miller, 1984), and showed greater awareness of commonalities between different groups (Gaertner & Dovidio, 2000; Tam et al., 2007). However, there is also evidence suggesting that such interventions may have minimal positive effects, no effects, or even negative effects on reducing personal bias and improving inter-group relations (Hite & Mc Donald, 2006; Kalev, Dobbin, & Kelly, 2006; Stewart, Latu, Branscombe, & Denney, 2010). Moreover, recent research shows that having a strong focus on positive contact underestimates the effect of negative contact, which has been shown to dull the effects of positive contact on prejudice (Barlow et al., 2012).

Another major form of prejudice-reduction intervention is education, which is based on the assumption that prejudice is the result of the individual's distorted thinking about out-group members due to their exposure to biased information in the environment, for example, in mass media (Birtel & Crisp, 2015). To reduce prejudice, education-based interventions focus on replacing biased information with accurate knowledge. In previous studies, researchers have designed various education-based approaches, such as presenting participants with print, audio, or visual materials that portray no ethnic or racial bias (Bigler & Liben, 2007; Verkuyten & De Wolf, 2007), providing moral rationale for not using prejudiced language (Aboud & Miller, 2007), and implementing multicultural curricula (Perkins & Mebert, 2005).

Compared with contact-based interventions, education-based interventions have been shown to produce more positive effects on changing individuals' prejudice (Aboud et al., 2012), and this has consequently attracted the attention of social workers. One recent example is the *HEAR.US Project*, supported by Tacoma Community House, which was designed to replace myths about immigrants with accurate knowledge about immigrants' experience in the United States (Epps, 2015). In this project, social workers gathered stories from immigration detention officers who witnessed immigrant families being torn apart under the enforcement of current immigration policies, and facts about immigration that clarify the negative portrayals of the immigrant group, for example, that they are draining resources and taking away jobs.

The overarching goal of contact and education is to develop positive attitudes toward out-group members and to increase mutual respect and understanding among individuals from different ethnic groups. Although individuals' positive attitudes may be created through these interventions, these forms of contact and education do not improve policy orientations of immigrants and societal normative views of immigrants. Promoting positive intergroup interactions, however, may intensify disadvantaged group members' perceptions of widespread unjust treatment, and result in the acceptance of a permanent disadvantaged status quo (Cakal, Hewstone, Schwär, & Heath, 2011; Saguy, Tausch, Dovidio, & Pratto, 2009). To reduce prejudice and discrimination, interventions may need to focus on mobilizing the disadvantaged to challenge the status quo (Dixon, Durrheim, Stevenson, & Cakal, 2016) and change societal norms about the expression of prejudice against immigrants.

2.3.2 Protest

Protest is a form of sociopolitical collective action in which members of a group act together to express objection to particular actions or situations (Amenta & Young, 1999; Burstein, 1999). It can take many forms, such as letter writing, public denunciations, marches, sit-ins, and boycotts directed toward prejudiced, offensive or stigmatizing practices (Corrigan, Roe, & Tsang, 2011). Using protest as an intervention strategy originated from research concerning public stigma against persons with mental illness. Given the increased coalescence and cross-communication of stigma and prejudice research (Phelan, Link, & Dovidio, 2008), the interventions that reduced stigmatizing public statements about mental illness have important implications for prejudice-reduction. This section reviews current research on protest against stigma and prejudice to highlight the need for examining the role of protest in prejudice against immigrants.

Protest is often applied against stigmatizing public statements, media reports, and advertisement (Rüsch et al., 2005). In the protest approach to stigma reduction, individuals or organizations attempt to change social norms by explicitly discouraging the stigmatization of a certain group (Casados, 2017). When discussing the important role of protest against stigma, previous researchers have cited anecdotal evidence to suggest that protest can reduce harmful media representations (Casados, 2017; Corrigan et al., 2012; Corrigan & Watson, 2002; Rüsch et al., 2005; Wahl, 1997). One of the widely cited protests was the National Alliance on Mental Illness (NAMI) campaign that aimed to remove stigmatizing depictions of mental illness in media. The protest took various forms, including writing letters criticizing TV episodes that depict people with mental illness as violent or offensively comedic, and encouraging the use of individual protest through publicly declaring themselves as #IAmStigmafree online (Casados, 2017). Anecdotally, these strategies may have had some positive effects on suppressing stigmatized attitudes. For example, NAMI initiated letter-writing campaigns in 2000 to protest against the television drama *Wonderland* because the show portrayed people with mental illness as violent. Immediately after the protest, the American Broadcasting Company pulled *Wonderland* after two episodes (NAMI, 2000). The success of campaigns might also be due to the effectiveness of protest in getting stigmatizing images of mental illness withdrawn from public conversations (Corrigan & Watson, 2002). Wahl (1997) suggested that because of this function of protest, citizens would be exposed to far fewer sanctioned examples of stigma and stereotypes, and consequently express fewer stigmatized and stereotypical comments about people of mental illness.

While anecdotal evidence suggests that protest can be an effective anti-stigma approach, only a few studies have tested its effectiveness. In a meta-analysis, Corrigan et al. (2012) examined publications between 1972 to 2010 that focused on the effects of the anti-stigma approaches on public stigma related to mental illness. Among 72 examined studies, only one tested the effectiveness of protest, which yielded non-significant findings for the effect of fact sheets from Psychiatrists' Changing Minds campaign on reducing stigmatized attitudes against schizophrenia and alcoholism (Luty, Umoh, Sessay, & Sarkhel, 2007). The results of the meta-analysis revealed that while there was abundant evidence for suggesting the effectiveness of education and contact on improving attitudes, affect, and behavior intentions toward individuals with mental illness, evidence regarding the effects of protest against public stigma was rare.

To further understand the effectiveness of different anti-stigma approaches, a recent study compared the effects of contact, education, and protest on attitudes, emotions, and behavioral intentions towards people who stutter (Boyle et al., 2016). The researchers randomly assigned participants to a control condition and one of three anti-stigma conditions. Findings of this study showed consistent evidence for the positive role of education and contact in reducing stigma. Moreover, protest was found to have a positive effect on reducing negative stereotypes, and its positive effect was maintained at one-week follow-up. This study suggested that protest could be an effective anti-stigma strategy to reduce negative attitudes about people with mental illness.

In prejudice research, there has also been increased attention to using protest as a prejudice-reduction intervention strategy in recent years. One major motivation is that prejudice-reduction interventions aiming to promote positive inter-group interactions have been shown to decrease perceptions of inequality and support for the implementation of social change among historically disadvantaged groups such as women and racial minority groups (Dixon, Tropp, Durrheim, & Tredoux, 2010; Ellemers & Barreto, 2009; Saguy et al., 2009). To address this issue, Dixon, Levine, Reicher, and Durrheim (2012) suggested a collective action model of prejudice reduction should be used in changing prejudice embedded in social structure.

According to this model, dominant group members are ranked higher in the hierarchy and rarely give away their power and privileges. To maintain group-based hierarchy and in-group dominance, dominant group members express prejudices against subordinate group members to minimize group conflicts and create consensus on ideologies that promote the superiority of one group over others. To change prejudice embedded in the group-based hierarchy, subordinate groups must collaborate to challenge the dominant group's advantage and work toward changing social norms that sustain the expression of prejudice. Findings from a recent study supported this premise (Acar & Uluğ, 2016). The 2013 Gezi Park protests initially aimed to contest the urban development plan for Istanbul's Taksim Gezi Park, but ultimately the goal that developed was to bring an end to police violence. The researchers found that protest participants achieved a common ground that improved the status of all present, participants reporting, for instance, that they overcome past prejudices against other disadvantaged groups (e.g., LGBT groups) during the protests. While this study suggests that participating in a protest can help protest participants overcome their previous prejudices, the impact of protest on non-participants' prejudices remains unclear.

A survey of previous research on prejudice-reduction shows that protest can be used as a strategy to resist prejudice and discrimination. However, whether protest can succeed in reducing prejudice still remains unknown. Therefore, it is necessary to survey the literature beyond the current prejudice and stigma research.

2.3.3 Exposure to protest and attitude change

Participants in a protest often believe that their actions can make the public more aware of certain critical issues (Giugni, 2004), pressure the government to take actions in policy change (Amenta & Young, 1999; Burstein, 1999), and change social values and norms (Banaszak & Ondercin, 2016; Faderman, 1992). Recent research suggests that protests could serve as critical counter-political voices (Tyler & Marciniak, 2013) for immigrants to resist prejudice and discrimination (Verkuyten & Martinovic, 2015). A large number of people are exposed to these immigrant protests either by physical presence at protest scenes or via media reports. The following section surveys current research on the role of exposure to protest in attitude change.

Protest can generate attitude change by invoking a broader identity that embraces both protesters and bystanders. By engaging in nonviolent resistance such as sit-ins, rallies, and boycotts, protesters provide information about their demands and their willingness to bargain (Chenoweth, Stephan, & Stephan, 2011), thereby invoking a common identity. The resultant dismissal, harassment, and even mass arrests are clear indications to the broader public of the social price that protesters are willing to pay, and bystanders who witness these costly actions may become more sympathetic to them. In past work that focused on the effects of protest on bystanders and the public at large, researchers have found that immigrants' rights protests in the US not only influenced Latinos' sense of empowerment and alienation (Wallace, Zepeda-Millán, & Jones-Correa, 2014), and their awareness of immigration related issues (Carey Jr et al., 2014), but also their group identity (Silber Mohamed, 2013). After 2006 immigrants' rights protests, Latinos, particularly Mexicans and Dominicans, were more likely to identify themselves as Americans. These findings suggest that protest can be effective at invoking broader identities, which can in turn reduce people's prejudice.

Exposure to protest movements may not change attitudes overnight, but it can lead to attitude change over time. In a study of the US women's movement from 1960 to 1962, Banaszak and Ondercin (2016) found that citizens adopted more liberal gender attitudes[2] as the movement increased its protest activities over time. Using historical data on US civil rights protests during

2 Opinions and attitudes are used interchangeably in this dissertation. Political science tends to favor the term opinions, while social psychology tends to favor attitudes. These two terms have been suggested using as synonymous in a theoretical discussion of these two constructs (Bergman, 1998) for the reason that distinguishing attitudes from opinions often contributes to a confusion rather a clarification (McGuire, 1985).

1960-65 with contemporary public opinion data, a recent study found that white Americans from counties that experienced civil rights protests tended to display more positive attitudes towards racial minorities compared with white Americans from counties that did not experience protests (Mazumder, 2018). The findings of these studies are consistent with a previous study (Boyle et al., 2016) on the effect of protest on stigma reduction in the experimental setting, where protest was found to have a positive effect on reducing negative stereotypes, and the positive effect remained at one-week follow-up. Together, these studies suggest that protest can improve gender and racial attitudes over time, and the impact of protest on people's attitude change can persists after the protest activities end.

Spatial proximity to political protests can also shape people's attitudes towards a group or related social issues. Wallace et al. (2014) found that during the 2006 immigrant protests, Latinos located within 100 miles of multiple small marches (fewer than 10,000 participants) reported more positive attitudes toward trust in government and stronger feelings about their agency in bringing about political and social change, whereas proximity to large-scale marches was associated with lower feelings of efficacy. Residing in the counties where there were high levels of immigrant protest activities was found to be a factor in Latinos' attitudes towards immigration policy (Branton, Martinez-Ebers, Carey, & Matsubayashi, 2015). White Southerners living in a county where a sit-in occurred were more likely to support the sit-in, whereas this effect was not observed for counties where a sit-in did not occur (Andrews, Beyerlein, & Tucker Farnum, 2015).

Recent research has also found that the impact of temporal and spatial exposure to protest on attitude change differs by individual's group identity (Branton et al., 2015). After the 2006 immigration protests, while foreign-born Latinos reported more positive attitudes towards immigrants and support for benign immigration policy (e.g., immediate legalization of current unauthorized immigrants), US-born Latinos reported less positive attitudes and support for the policy.

These studies are the major building blocks for understanding the role of exposure to protest in prejudice against immigrants, because their findings suggest that temporal and spatial exposure to protest can change people's attitudes. However, it still remains unclear whether exposure to protest can change prejudice against immigrants. Given that prejudice has been theorized as attitudes (usually negative) towards a group or members of a group, it should follow that exposure to protest can alter people's prejudice against immigrants. But because people with a different social identity may respond to protest differently as suggested by Branton et al. (2015), it is also important to consider the role of social identity in altering the relationship between protest exposure and prejudice against immigrants. Therefore, this dissertation takes the first step to examine the role of exposure to protest in people's prejudice against immigrants and how social identity alters the relationship between the two.

2.4 EMERGING ROLE OF SOCIAL MEDIA IN PREJUDICE

Today, seven-in-ten Americans use social media to connect with one another, engage in political conversation, share information; many also use it to promote protest and collective action (Pew Research Center, 2018). As we are living in an increasingly online society, the role social media plays in prejudiced speech and interventions that intend to address the problem should not be overlooked. Previous research has suggested that social media is a double-edged sword in addressing the issue of prejudiced speech. While it gives rise to profoundly unsettling forms of, and intensification of, prejudiced speech, it also provides unprecedented tools to coordinate massive and rapid public responses to these issues. The following section provides an overview of the negative and positive roles of social media in prejudice to highlight the importance of examining online prejudiced speech.

2.4.1 Negative role of social media in prejudice

Social media provides a platform for the public to express or share information openly and freely with almost no filtering (Sayre, Bode, Shah, Wilcox, & Shah, 2010), its emergence largely changing the dynamics of how people communicate. Compared to face-to-face communication, computer-mediated communication lacks the social restrictions and inhibitions that prevent people from speaking their prejudices in public (Lapidot-Lefler & Barak, 2012; Suler, 2004). As a consequence, social media creates a public space where people think it is socially acceptable to openly express prejudices of any kind (Spata, 2015). While previous research has long been concerned with the corrosive effects of prejudice, the advent of social media brings new urgency to this concern.

Recent studies show that the adverse effects of expressions of prejudice are as real in online social media as they are in so-called "real world" offline settings (Kowalski et al., 2014). Online verbal attacks have been linked to adverse psychological outcomes such as depression, anxiety, and lower self-esteem (Kowalski et al., 2014). People from minority ethnic, gender, and sexual minority groups have continued to be the primary targets for various forms of prejudice in social media (Rubin & McClelland, 2015). Indeed, immigrants may be particularly susceptible to the real, adverse effects of online prejudice. As shown in a recent study, immigrant youths are more likely than US-born youths to become victims of online verbal aggression (Comas-Forgas et al., 2017).

Moreover, prejudice in social media is often more overt and extreme than what is encountered offline (Burnap & Williams, 2016; Suler, 2004). One of the extreme forms of prejudice is hate speech. There has been growing concern about the growth of online hate groups and their influence on shaping social values and perceptions (Chau & Xu, 2007; Lee & Leets, 2002; McNamee, Peterson, & Peña, 2010). Hate groups such as the KKK, Neo-Confederate, White Supremacists, and Black Separatists have used social media as a tool to spread racist beliefs and to incite violence offline (Ben-David & Matamoros-Fernandez, 2016; Chan et al., 2016). This is of particular concern for youths and young adults because they are among the earliest social media adapters, who are now, and will likely continue to be, a major component of social media users (Greenwood, Perrin, & Duggan, 2016). Moreover, young people are more likely to be affected and persuaded by hatred and extremist ideas propagated through the Web (Chau & Xu, 2007). Expressions of prejudice and even hatred may become social norms in social media as hate groups become more prevalent and influential online.

Several studies have examined the dynamics of hate speech, and found that offline conflict events, such as terrorist attacks, tend to amplify hate speech in social media (Burnap & Williams, 2014; Burnap & Williams, 2015; Magdy, Darwish, & Abokhodair, 2015; Williams & Burnap, 2015). For example, Burnap and Williams (2015) found that following the London Bombing of July 7, 2005, there was an increase in using “othering” words in hate speech against Muslims to express the intention to exclude Muslims (e.g., “send them home”), to justify their expectations of malicious behavior from Muslims (e.g., “told you so”), and to openly disparage Muslims (e.g., “Muslim savages”).

A few recent studies also provide insight into the demographics of hate speech perpetrators in social media. Waseem and Hovy (2016) showed that men are more likely than women to express racist and sexist hate speech on social media, and despite the differences in the two 2016 US Presidential candidates, another recent study found that followers of Hillary Clinton and Donald Trump were equally likely to make racist comments (Lozano et al., 2017).

Studies have also examined the influence of online hate speech. In discussions of unauthorized immigrants, immigration policy, and border security, Twitter users expressing negative emotions such as fear, disgust, anger, and distrust tend to be more influential than those expressing positive and neutral tones (Chung, He, Zeng, & Benjamin, 2015). This asymmetry suggests that being more influential, prejudiced users play an important role in promoting prejudice toward immigrants; this has important implications for the design of online interventions.

In addition to research on the dynamics and effects of online hate speech, there is substantial literature on the purely technical means of detecting and measuring hate speech (Awan, 2014; Badjatiya, Gupta, Gupta, & Varma, 2017; Burnap & Williams, 2016; Gitari, Zuping, Damien, & Long, 2015; Kwok & Wang, 2013). Beyond hate speech, there have also been studies on rumor spreading (Starbird, 2016; Zeng, 2016), age stereotypes (Levy, Chung, Bedford, & Navrazhina, 2013), Islamophobia (Awan, 2014), verbal aggression (Chatzakou et al., 2017; Golbeck et al., 2017; Guberman, Schmitz, & Hemphill, 2016), and group risk perception (Chung, Wei, Yu, & Wen, 2016). These studies showcase the corrosive effects of social media on spreading hate and prejudiced language. The lack of understanding of its dynamics may limit and even produce counterproductive effects of protest and collective action.

2.4.2 Positive role of social media in prejudice

While online social media can intensify prejudiced speech, it also provides social workers with new tools and unprecedented opportunities for scaling up advocacy. One of the major advantages, as mentioned previously, is that social media enables massive, rapid, and spontaneous public responses to critical issues (Shirky, 2011). According to previous studies, social media is also particularly well suited to promoting offline collective action (De Choudhury, Jhaver, Sugar, & Weber, 2016; Gil de Zúñiga, Jung, & Valenzuela, 2012; Valenzuela, 2013; Zhang, Johnson, Seltzer, & Bichard, 2010).

Bringing public attention to bear on issues of the moment is a necessary step for promoting social change, but there is evidence that using social media to share personal experiences can also transform public perceptions of marginalized groups such as lesbian, gay, bisexual, and transgender people (Jones, 2015). For example, the *It Gets Better* project, a successful campaign that transformed public perceptions of LGBT groups (Jones, 2015), was initiated in response to the publicized suicide of Bill Lucas, a teenager who hung himself after suffering anti-gay bullying. Jones (2015) found that the success of the campaign was largely dependent on the stories that participants told their listeners. Their stories (in the form of YouTube videos), tailored to their audiences, made it possible to comfort the victims and confront the perpetrators.

Previous researchers also found that the social media campaign centered on #BringBackOurGirls sparked a collective global response. Twitter users in the United States used this hashtag to demand not only the release of the Nigerian schoolgirls kidnapped by a terrorist organization, but also the rights of formal education for women in Nigeria (Chiluwa & Ifukor, 2015). During the Black Lives Matter protests, research showed that bystanders used #Ferguson to provide social support to participants (Bonilla & Rosa, 2015). More recently, the hashtag

#MuslimsAreNotTerrorists was used among Twitter users worldwide to defend Muslims and to reject the Muslim ban proposed in the United States after the 2015 Paris attacks (Magdy et al., 2015; Moon & Fares, 2016).

People who use social media to express opinions have been shown to be more likely to attend public demonstrations (Valenzuela, 2013), and moreover those who use social media frequently are more likely to attend protests (Gil de Zúñiga et al., 2012; Park, Kee, & Valenzuela, 2009; Rojas & Puig-i-Abril, 2009; Valenzuela, Park, & Kee, 2009; Zhang et al., 2010). Increased social media activity and engagement (e.g., number of posts shared, number of retweets, etc.) are related to an increase in participants in an offline protest (De Choudhury et al., 2016). These studies have provided important insights for the role of social media in promoting offline protests (Chiluwa & Ifukor, 2015) and raising awareness among the public (Jones, 2015). However, few studies have examined the impact of offline protests on changing people's prejudicial speech on social media.

In summary, the same social media tools that are used to spread prejudiced speech and political agendas can and must be used to advocate for social change on behalf of immigrants, and empower immigrants to advocate for themselves. However, since the dynamics of online social interaction are scarcely understood, there is a significant risk of unintended consequences arising from online promotion of social protests, for example, provoking a backlash (Wang & Piazza, 2016). To make effective use of social media, more needs to be known about the dynamics of prejudiced speech online, and the interplay of online prejudice with offline social protests. Therefore, this dissertation takes the first step to examine the impact of offline immigrant protests on prejudice toward immigrants in social media.

3.0 THEORETICAL FRAMEWORK

In the past several decades, researchers have expressed a strong interest in understanding prejudice. This dissertation utilizes three theories to explain the role of protest in prejudice: social movement impact theory, a theory of civil resistance and long-run attitude change, and intergroup emotions theory. This section reviews the key concepts and assumptions of each theory, and provides a synthesis of the empirical evidence that supports these theories.

3.1 SOCIAL MOVEMENT IMPACT THEORY

Social movement impact theory is useful for assessing the impact of protests on society. According to this theory, protests can lead to changes in political and policy outcomes (Gamson, 1975; Piven & Cloward, 1977) and can influence cultural outcomes (Bernstein, 2003). The following sections describe the premises of social movement impact theory and empirical evidence that supports the theory.

3.1.1 Key concepts and assumptions

The fundamental assumption of social movement impact theory is that protest can have an impact on society. Gamson (1975) first introduced this theory in his book *The Strategy of Social Protest*. This work is one of the most comprehensive studies that identified the characteristics that distinguish successful protests from their unsuccessful counterparts. In this work, Gamson analyzed a random sample of 53 American protest groups (e.g., Christian peace movements) from a list of 4,500 groups that participated in social protest between 1800 and 1945. He found that the characteristics of protest groups such as tactics and goals influence their chances for success in promoting political and policy changes. For example, protest groups that used disruptive tactics such as strikes were more likely to succeed in promoting political and policy changes than those did not use these tactics. This finding is important to this research because disruptive tactics such as a strike was one of tactics used in the “Day Without Immigrants” protest.

Piven and Cloward (1977) further developed this theory by examining protests during the Great Depression by low-wage and unemployed industrial workers, and protests following World War II by African Americans in southern US states. Their findings upheld Gamson’s (1975) conclusion that protest was effective in creating political and policy change. Further, they pointed out that the success of a protest group depended not only on their goals and tactics, but also on the social context, for example, the relative unity or disunity of active opponents of the protest group as well as the perceptions of that part of the public which is, if not neutral at least can be persuaded to support the goals of the protest. In the context of online space, however, it is unclear if the social basis of the protest group's success still holds, where social media platforms have created “filter bubbles” that amplify the unity or disunity of active opponents of an issue. (Bakshy, Messing, & Adamic, 2015; Pariser, 2011).

Social movement impact theory also poses that protest has an impact on cultural outcomes, for example, changing public understandings and belief systems that have a bearing on a protest issue (Bernstein, 2003; McAdam, 2010). Previous research identified public opinion as indicators of cultural outcomes. This is because public opinion represents, what most people think about particular issues, i.e., a normative response to those issues (Glynn, 1997). Protest can change the way people think about a particular issue. For example, women suffragist movements in the US changed the way Americans think about gender roles (McCammon et al, 2008). In addition to public opinion, Bernstein (2003) also identified the discursive impact of protest as an important cultural outcome. Public discourse influences the way bystanders understand the issues illuminated by the protesters. Public discourse influences the way bystanders understand the issues illuminated by the protesters. This is because protesters bring credibility to their claims, and cast doubt on the accepted truth, thereby transforming the discourse in such a way as to achieve an equal, if not dominant, voice (Steinberg, 2002). To achieve desirable outcomes, Bernstein (2003) suggested that protesters use appropriate rhetoric and formulate their message so that it is effective and appealing to bystanders. By creating and shifting discourses, protest can alter the ways issues are understood and transform public understanding and belief systems on an issue (Jacobsson & Lindblom, 2016). These theoretical claims are important for this research because they highlight the impact of protest on public discourse. Just as online prejudiced speech can be understood as a form of online discourse, so it is possible that protest can influence online prejudiced speech.

3.1.2 Empirical evidence

Past research into social movement theory has provided substantive evidence for the impact of protest on our society. The following section presents the major empirical evidence that supports this assumption.

Research examining movement outcomes primarily focused on the extent to which social protest and movements has led to political and policy changes. For example, protests that oppose the United States' involvement in the Vietnam War were found to be positive related to a shift in anti-war voting patterns in the Senate (Burstein & Freudenburg, 1978). The civil rights movements was found to impact voting patterns such as number of Black voters registered and the number of Black elected officials (Andrews, 1997). The Tea Party Movement in the United States in 2009 increased public support for Tea Party positions and led to more Republican votes in the 2010 midterm elections (Madestam, Shoag, Veuger, & Yanagizawa-Drott, 2013). A recent study that examined the impact of the 2006 immigration protests on Latinos' immigration policy preferences found that that foreign-born Latinos surveyed after the protest showed more support to amnesty (immediate legalization of current unauthorized immigrants) than they did before the protests (Branton et al., 2015). Together, these above studies uphold the assumption that social protest and movements can an impact on political and policy outcomes.

While research examining movement outcomes primarily focused on its impact on changes in political outcomes, a few studies also started to explore its impact on cultural outcomes. Research examining the impact of protests on public opinion suggests that rather than directly impact political and policy change, protests and social movements that affect policymaking by influencing public opinion (Banaszak & Ondercin, 2016; Branton et al., 2015; Madestam et al., 2013). For example, the US women's movement from 1960 to 1962 introduced a more liberal view of women's roles and status in family and work (Banaszak & Ondercin, 2016). McCammon et al. (2008) found that in the women suffragist movements, suffragists were more successful in convincing lawmakers and the public to give women voting rights when they used arguments that tapped into beliefs widely accepted at the time such as women would use the vote to protect children and the home. In a more recent study, Mkono (2018) found that the Cecil anti-trophy hunting movement led to a discussion of ethical ramifications related to trophy hunting on social media. The Cecil anti-trophy hunting movement was a response to trophy hunting in which Cecil, the Lion that lived in a Zimbabwe national park, was shot in 2015 by the American tourist, Walter

Pal mer. This movement sparked a global “cybermovement” against trophy hunting. In the discussion of this issue, trophy hunters were framed by the public as “murderers” who deserved the same measures of violence’s as they inflicted on animals. For example, people posted their comments in Youtube, “I don’t understad people who hunt for hobbie. Whats the point of killing an animal that you will not eat? Its just as bad as killing a human” (p. 1615). This framing of trophy hunters helped promote the protestor’s goal to prohibit trophy hunting. This study suggests highlighted the importance of using social media to shape social discourse in achieving the goal of a protest. However, it still remains unknown about whether protest can influence online discourse such as prejudiced speech.

In summary, while past research has provided strong evidence for supporting the impact of social movement on political outcomes, only few studies examined its impact on cultural outcomes. It is also worth noting that past researchers also have critiques about this theory such as problematic definitions of protest success and the issues of causality. As related to this dissertation, online prejudiced speech can be understood as an indicator of cultural outcomes because it taps into public discourse occurring in online space. However, it remains unknown whether protest can influence online discourse such as online prejudiced speech.

3.2 A THEORY OF CIVIL RESISTANCE AND LONG-RUN ATTITUDE CHANGE

Mazumder (2018) developed a theory of civil resistance and long-run attitude change to explain why nonviolent protest can shift attitudes and why the changes in attitudes can persist, proposing that instances of collective action by protest movements can generate attitude change among the target public. In addition, long after a social movement's activity, historical ideational change persists through a system of intergenerational socialization. Together, these premises suggest social movements can give rise to changes in attitudes towards a group in the long run, which makes this theory well-suited to explain the role of protest in reducing prejudice because prejudice is a kind of attitude, and the major evidence supporting this theory shows the ability of protests to reduce racial prejudice. The following section provides the case used to develop the theory, its premises, and the empirical evidence that supports it.

3.2.1 Key concepts and assumptions

The case of the American civil rights movement was used to provide theoretical and empirical leverage in the development of the theory of civil resistance and long-run attitude change. The civil rights movement was a turning point in US history, which makes it relevant for understanding the long-term impact of historical social movements. More importantly, this case was concerned with the fundamental reshaping of American values about racial attitudes. In the context of the American civil rights movement, racial relationship was primarily concerned with the relationship between whites and blacks. The American establishment was founded on a notion of the superiority of whites over blacks (King & Smith, 2005), a white supremacist ideology has been infused in the institutions that maintained the hierarchy

Mazumder (2018) argued that protest can impact prejudice. This is assumption is built on previous studies on the impact of social protest and movements on public opinions and literature examining psychological and informational mechanisms. Here, I primarily discuss studies that focus on psychological and informational mechanisms because previous section on social movement impact has discussed the impact of social protest and movements on public opinions extensively (§3.1).

The ways in which protests affects whites' attitudes towards blacks can be explained the effect of protest on priming social identities that exist beyond race. The idea of priming identity is built on work from ingroup identity model proposed by Gaertner and Dovidio (2000) . In an early experiment, Gaertner et al. (1989) showed that inducing participants to recategorize their ingroup identity into one common group identity instead of multiple groups reduced their prejudice against former outgroup members. Similarly, experimentally manipulating white respondents to feel closer their American identity was found to increase their support to fund public programs that would help racial minorities (Transue, 2007) . Built on these previous work on the priming effect of social identity, Mazumder (2018) argued that exposure to local civil right protests could led whites to feel closer to blacks because the protest highlighted more transcendent identities and the ways in which whites and blacks are connected, thereby reducing whites' prejudice against African Americans. Attitude change then persists through intergenerational socialization, process through which parents have incentives to inculcate their children with cultures and attitudes similar to their own (Boyd & Richerson, 2005; Tabellini, 2008). This psychological mechanism of priming effect of social identity and intergenerational socialization together could explain the long-lasting effect of civil rights movements on prejudice change.

In addition to psychological mechanisms, Mazumder (2018) also argued that the impact of protest on prejudice could come from persuasion and informational channels through which protest influences political attitudes and public opinions. To support this argument, he cited studies that examined the effect of persuasion on changing people's attitudes. For example, Broockman and Kalla (2016) found that a conversation related to discrimination against transgender persons led participants to be more sympathetic toward transgender rights and that these effects persist for a significant amount of time. In this case, protesters could invoke a conversation and provide a more vivid understanding about oppression encountered by minorities, which in turn might lead attitude change. This is because by engaging in activities such as protest rallies and boycotts, protesters provide information about their demands and open them to repression. For example, protestors in the civil rights movements, remained peaceful despite being arrested and beaten and harassed by local authorities and hate groups. These might lead to the bystanders to become more sympathetic to protestor's demands and therefore influence their attitudes.

3.2.2 Empirical evidence

To test the theory of civil resistance and long-run attitude change, Mazumder (2018) obtained civil rights protests data, using dynamics of Collective Action data set, which records demonstrations of collective action from 1960 to 1965. Pool survey data from the Cooperative Congressional Election Study (CCES) was used in measuring white's racial prejudice against blacks. This data contains over 157,000 white respondents from 2006 to 2011 and their self-reported prejudice against African Americans. For example, white respondents were asked to indicate how strongly they agree, or disagree with the statement such as, "The Irish, Italians, Jews and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors".

The protest and prejudice data were aggregated to county level to capture the local impact of civil rights protest on whites' prejudice against blacks. Findings of this study show that at county level, whites from counties that experienced civil rights protest were less likely to express racial prejudice against blacks compared to whites from counties that did not experience the protest (Mazumder, 2018).

Given that this theory has been proposed just recently, it has not been tested substantially. Future discussion of this limitation is included in section 6.2. Nonetheless, evidence from the impact of exposure to protest on attitude change can lend support to this theory, where past research has shown that exposure to protests across a variety of areas, such as issues related to mental illness, gender, and immigration, were associated with improvement in attitudes toward these issue-relevant groups (Banaszak & Ondercin, 2016; Boyle et al., 2016; Branton et al., 2015). For example, foreign-born Latinos adopted more positive attitudes about undocumented immigrants after exposure to the 2006 immigrant protest (Branton et al., 2015). The overall implication from these studies is that protest can affect attitudes and potentially reduce prejudice.

3.3 INTERGROUP EMOTIONS THEORY

Eliot Smith and Diane Mackie developed Intergroup Emotions Theory (IET). According to IET, people possess multiple social identities that reflect their self-identified membership in various social groups; a person's social identity can vary in place and time (Mackie, Maitner, & Smith, 2009; Ray, Mackie, & Smith, 2014). The way a person appraises a group and its actions will depend on which of these social identities is most salient at any given moment. Different appraisals can, in turn, give rise to different positive emotions (admiration, sympathy, etc.) and negative emotions (hatred, anger, fear, etc.).

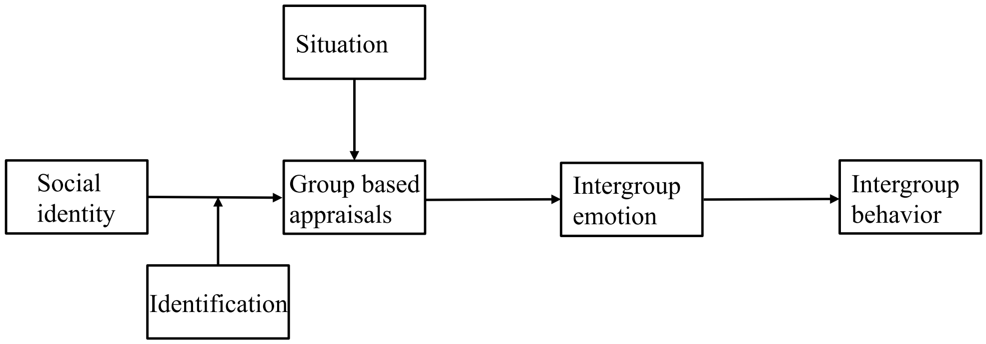

Figure 1 A model of intergroup emotion theory

Figure 1 shows the model of intergroup emotion. According to IET, when people's social identities become salient, they appraise ongoing events and entities through an in-group lens and respond to them with corresponding in-group emotions (Smith & Mackie, 2015). Therefore, when an event negatively influences a group, people identified with that group will interpret the event negatively and have negative emotional reactions toward the event, even if they are not directly harmed. In this process, stronger identification with the in-group leads to more intense emotional reactions. The following section provides the premises of this theory and empirical evidence that supports it.

3.3.1 Key concepts and assumptions

IET has two fundamental assumptions. First, a person possesses multiple social identities that can vary in time and place, and that while these social identities are distinct from a person's more or less invariant personal identity as a unique individual, they play an analogous role ("We" in place of "I") as a locus used by the self to interpret events in the world. Second, the phenomenology of social emotions arising from social identities is related to but distinct from the phenomenology of individual emotions arising from individual identities.

Social identity in IET builds on social identity theory (Tajfel & Turner, 1986) and self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). In both theories, people form their social identities when they shift from seeing themselves as unique individuals to seeing themselves as interchangeable exemplars of a social group (Smith, 1993; Turner et al., 1987). This process from "me" as an individual to "us" as a member of an in-group is accompanied with a parallel shift from "you" as an individual to "them" as a member of an out-group (Ray et al., 2014). As a result, individuals view the world through a group lens and acquire emotions that are reinforced by other self-defined members of that group. People possess social identities that are relatively stable, for example, those that reflect religious and professional affiliations; social identities can be made salient by situations that highlight inter-group conflicts (e.g., terrorist attacks and protests).

Social emotion in IET builds on appraisal theories of emotions (Lazarus, 1991; Smith, 1993), which claim that emotions arise from the appraisal of events by a person (in the structure of IET, from the appraisal of events from the vantage point of a person's individual identity). IET extends this construct to claim that social emotions arise from the inter-group appraisal of events attributed to some group, by a person's salient social identity (Mackie et al., 2009). The role of social groups in the observer's social identity and in the attribution of events makes intergroup appraisal in IET quite distinct from individual appraisal in appraisal theories of emotion. While emotions are always felt by an individual (the appraiser) the term "social emotion," however, is used to distinguish emotions arising from group-based rather than individual-based appraisals.

3.3.2 Empirical evidence

Past research into intergroup emotions theory has provided substantive evidence for the relationship between intergroup emotions and people's social identities (Seger, Smith, & Mackie, 2009) and behavior intentions (Leonard, Mackie, & Smith, 2011; Leonard, Moons, Mackie, & Smith, 2011; Maitner, Mackie, & Smith, 2006). Previous evidence showed that intergroup emotions theory provides a sound framework for understanding nuances in prejudice, and examining its situation-specific and episodic patterns (Miller, Smith, & Mackie, 2004; Ray et al., 2008; Ray, Mackie, Smith, & Terman, 2012; Rydell et al., 2008). The following sections present the major empirical evidence that supports the conceptualization of prejudice as emotional, and the relationship between social identity, intergroup appraisals and inter-group emotions.

Evidence has consistently illuminated the positive relationship between group-based anxiety and prejudice (Britt, Bonieci, Vescio, Biernat, & Brown, 1996; Islam & Hewstone, 1993; Murray & Marx, 2013; Stephan, Ybarra, & Bachman, 1999). Dijker (1987) found a strong correlation between the attitudes of study participants toward ethnic minorities and the emotions participants experienced (e.g., anxiety and irritation) when coming into personal contact with members of those minority groups. Stephan and Stephan (1985) coined the term intergroup anxiety to describe people's feeling of anxiety when interacting with unfamiliar out-group members.

Cottrell and Neuberg (2005) also found that different out-groups could evoke different patterns of discrete emotions. They argued that generalized measures of prejudice, as positive and negative valences of attitudes towards an out-group, would mask a diversity of discrete emotional reactions. In the study, they found that while white American participants reported similar levels of negative attitudes toward African Americans and Native Americans, these resulted from different emotional reactions towards these groups. Negative attitudes of white American participants toward African Americans arose primarily from a combination of anger and fear, whereas the participants' negative attitudes towards Native Americans arose primarily from pity.

This diverse array of emotional reactions underlying attitudes toward out-groups was also found in a study that examined attitudes toward sexual orientation. Ray et al. (2012) found that straight men reported positive attitudes toward lesbian women and other straight men at similar levels, but the underlying emotional reactions of straight men towards these two groups were different. The positive attitudes of straight men towards lesbian women arose from emotions associated with feelings of sexual desire, while their positive attitudes to other straight men arose from emotions associated with camaraderie. These results showed the importance of understanding the emotional basis of prejudice, and supported the emotional conceptualization of prejudice.

In addition to evidence for intergroup emotion, previous studies examining the convergence and divergence of emotions within and between social groups provided strong evidence for the role of social identity in dictating intergroup emotions (Ray et al., 2014). Studies have shown that different emotional responses toward out-groups can arise depending on whether it is the individual identity or the social identity that is salient. When only social identities are salient, studies have also shown that a) participants can have different emotional responses to the same group depending on which of their social identities is salient; b) there is significant consistency in emotional responses among participants when the same social identities are salient.

Smith and colleagues (2007) studied the convergence and divergence of emotions among participants who were led to self-categorize themselves as individuals, Americans, and political party members. This study found that participants reported different emotions when they self-categorized themselves as members of a social group than they did as individuals. In addition, participants reported the same emotion when they self-categorized themselves as members of the same social group. A sense of pride was reported across all participants who self-categorized as Americans; and a feeling of anger was reported across all participants who self-categorized as Democrats. Further, these emotions could contribute to participants' group-relevant action beyond their individual emotions. For example, anger elicited by self-categorizing as a Democrat predicted the desire to argue with Republicans, whereas anger elicited by self-categorizing as an individual did not predict such desires.

Ray et al. (2008) further showed that self-categorization as members of different in-groups influenced their emotional reactions towards different out-groups. In the study, half of the American students were led to self-categorize as Americans and half as students. Participants reported more anger and less respect to Muslims when categorized as Americans than when categorized as students. In contrast, participants reported less anger and more respect to police when categorized as Americans than when categorized as students. In addition to these previous studies, a series of studies also demonstrated repeatedly that making salient one or another participant's social identity changed the emotional reactions to events (Dumont, Yzerbyt, Wigboldus, & Gordijn, 2003; Gordijn, Yzerbyt, Wigboldus, & Dumont, 2006; Tarrant, Dazeley, & Cottom, 2009; Yzerbyt, Dumont, Wigboldus, & Gordijn, 2003). These previous studies suggest that emotions felt about particular out-groups and events depend on which group the individuals identify with at that moment. Thus, in the context of intergroup conflict, different types of social identities dictate different emotional reactions toward an out-group.

Evidence also suggests that appraisals of inter-group threat mediate the effects of self-categorization on intergroup emotions (Kuppens, Yzerbyt, Dandache, Fischer, & van der Schalk, 2013; Smith & Mackie, 2015; Smith et al., 2007). Appraisals of intergroup threat refer to a person's appraisal of whether the out-group poses a threat to the in-group's rights and freedoms, physical safety, group values, or moral standings (Cottrell & Neuberg, 2003). Kuppens and Yzerbyt (2012) found that making participants' social identity as women the salient factor (neither as individuals nor students) led to an increase in fear, an effect that was mediated by appraisals of Muslims as a threat to physical safety. Moreover, participants' social identity as women also reported stronger intentions to avoid Muslims, an effect that was mediated by fear. These findings of the relationships among social identity, intergroup emotions and intergroup appraisals are consistent with the theoretical assumptions of the IET.

Overall, past experimental research has provided strong evidence for the IET's key premises. The emphasis on the role of emotions in prejudice enables the differentiation of emotional reactions directed toward different immigrant groups. For example, it allows this dissertation to explore the differences in emotional reactions toward immigrants in the context of intergroup conflict. Moreover, the conception of intergroup emotion as group-level emotion, and its relationship with social identity, provides new directions for analyzing and reducing prejudice. Investigating the divergence of emotions between social groups provides useful information for detecting the onset of prejudice, where interventions are most effective. The understanding of social identity in dictating intergroup emotions, such as changing the categorization of the self, provides insights for new strategies to reduce prejudice toward immigrants (Ray et al., 2008).

While the IET provides a sound framework for understanding prejudice, further development of this theory requires researchers to measure group-level emotions and appraisals repeatedly over time (Smith & Mackie, 2015). In addition, although the existing work on intergroup emotions has provided important insights on social categories such as national, ethnic, gender, and political party groups, little is known about the relationship between attitudinally defined groups (e.g., people who support or oppose a specific policy or action) and emotional reactions towards an out-group (Mackie & Smith, 2015). Moreover, whereas current research on the IET has focused on the negative impact of intergroup conflict events, the role those events played as a positive force promoting social change has been largely ignored.

4.0 METHODOLOGY

This dissertation adapted computational focus groups, a quasi-experimental research design, to examine the impact of protest on online prejudiced speech against immigrants. Specifically, I focused on two recent protests that targeted immigration issues: the "Day Without Immigrants" protest and the "No Ban, No Wall" protest. For each event, user timeline data were collected from Twitter. The sections below provide detailed descriptions of study design (see §4.1), protest events (see §4.2) and data collection (see §4.3).

4.1 STUDY DESIGN

Smith and Mackie (2015) suggest that new research approaches are needed to gain insight into the inter-group emotional dynamics of prejudice; in particular, approaches that (repeatedly and over time) measure group-level emotional reactions, and appraisals for events that trigger emotional episodes.

Computational focus groups, while not specifically used by inter-group emotion theory, is a methodological framework that satisfies the needs identified by Smith and Mackie (2015), albeit in a social media setting. Smith and Mackie identified this principal limitation of traditional approaches (i.e., laboratory experiments and survey) to studying prejudice: the difficulty of studying the time- and event-varying course of development and expression of prejudice. Adapting computational focus groups to the study of the group emotion component of prejudice towards immigrants overcomes this limitation. The following section describes the components of computational focus groups.

Computational focus groups is a framework for tracking changes in social media users' emotions, attitudes, or opinions about a group or an issue following specific events (Lin, Margolin, Keegan, & Lazer, 2013). Specifically, it tracks users' behavioral outcomes by analyzing the content of social media users' posts. This framework is similar to traditional intervention studies in that it requires an intervention (a focal event) and a measurable outcome (users' behavioral outcomes). However, the major difference between these two is the methods used to obtain outcomes. This is mainly because online users express their emotions, attitudes, and opinions in the form of unstructured texts, which requires researchers to leverage text mining techniques to turn unstructured texts into numbers.

In this dissertation, I adapt computational focus groups to study users' online prejudiced speech and how immigrant protest events are related to its changes. This framework consists of implementing the following steps (see Figure 2 below).

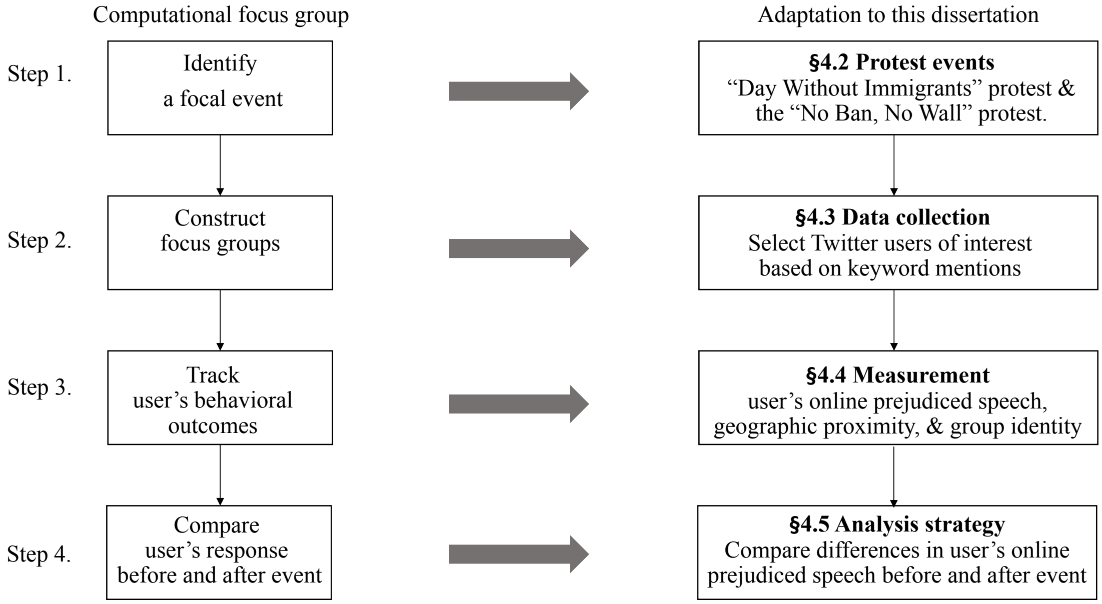

Figure 2 Four-step implementation of Computational focus group framework in this dissertation

Identify a focal event. A focal event is an event that has potential impact on people's behavioral outcomes. For example, previous research has used computational focus groups in studying events such as terrorist attacks and presidential debates (Lin, Keegan, Margolin, & Lazer, 2014; Lin, Margolin, & Wen, 2017). In this dissertation, I select two most recent immigrant protests as focal events: "Day Without Immigrants" protest and the "No Ban, No Wall" protest (see §4.2). These two events are selected because they are the most recent nationwide immigrant protest.

Construct focus groups. Focus groups, traditionally, are a form of group interview that capitalizes on communication between research participants in order to generate text data (Kitzinger, 1995). Social media users generate data by communicating their emotions, attitudes, or opinions about a group or an issue by posting short text messages. This wealth of data eliminates the need for group interviews. Instead, the computational focus groups framework adopts the concept of focus groups by examining the accumulated social media postings to understand users' behavioral outcomes. In this dissertation, I took the method of computational focus groups and constructed a user panel who showed interest in discussing the topics relevant to immigrants (see §4.3), and divided them into sub-groups based on their geo-proximity to protest cities (see §4.4.1) and self-reported group identity (see §4.4.2).

Track user's behavioral outcome(s). Users' behavioral outcomes are tracked by leveraging text mining techniques to quantify users' social media posts. Text mining relies on automated analysis of social media posts and is accomplished by building predictive models that identify features (in this case, number of words containing negative sentiment) extracted from social media data (Guntuku, Yaden, Kern, Ungar, & Eichstaedt, 2017). Past research that applied computational focus groups to track users' behavioral outcomes has leveraged text mining techniques to track user's group risk (Chung et al., 2016), and sense of distress (Wen & Lin, 2016). In this dissertation, I leverage text mining techniques to identify whether a tweet is prejudiced speech against immigrants or not; and whether a tweet is about immigrants or not. Then, I aggregate all the tweets to user-level (see §4.4.3).

Compare users' behavioral outcome(s) before and after event. In this framework, users' behavioral outcome(s) before a focal event is considered a baseline measure. The differences in the user's behavioral outcome(s) before and after the event are regarded as the changes related to the focal event. In this dissertation, I compared users' online prejudiced speech against immigrants before and after protest events (see §4.5).

4.2 PROTEST EVENTS

To understand the impact of protest on online prejudiced speech, this dissertation focused on the "Day Without Immigrants" protest and the "No Ban, No Wall" protest. These protest events were chosen because they were the most recent nationwide protests that aimed to show the important contributions of immigration and to resist punitive immigration policies. Background on these protest events is given below.

4.2.1 The "Day Without Immigrants" protest

The "Day Without Immigrants" protest was a response to President Donald Trump's plans to build a border wall, strip sanctuary cities of federal funding, and deport potentially millions of undocumented immigrants (Bermudez, Vives, Kohli, & Etehad, 2017). The protest took place on February 16, 2017 in multiple cities across the US, and was designed to demonstrate the importance of immigrants to the US economy and in the day to day lives of American citizens. Organized at the local grass roots level, people planned the action via social media (primarily Facebook and WhatsApp) and by word of mouth, discussing the action in restaurant staff meetings, on construction sites, and on commuter buses (Robbins & Correal, 2017; Welch & E, 2017).

Activists called for immigrants regardless of their legal status to stay home from work or school, close their businesses, and abstain from shopping. On the protest day, shops and restaurants were closed in several major US cities. For example, more than 50 restaurants were closed in Washington DC (Daniels & Graham, 2017), and over 1000 businesses were closed in Dallas (Reece, 2017). In addition to closed businesses, thousands of children did not attend school. For example, only 60% of students attended school in the KIPP Comunidad charter school network in Austin, Texas (Robbins & Correal, 2017), and only 65% of elementary students attended school in the Fort Worth, Texas Independent School District (CSB, 2017). To understand the impact of this protest, this dissertation focused on the role of protest exposure and social identity in online prejudice against immigrants.

4.2.2 The "No Ban, No Wall" protest

The "No Ban, No Wall" protest took place on January 28, 2017 as a response to President Donald Trump's plan to ban citizens of certain Muslim countries from entering the US, and suspend admission of all refugees entering the country (Demick, 2017).

As with "Day Without Immigrants," "No Ban, No Wall" was also planned via social media, and simultaneously executed in multiple cities in the US, including New York, Los Angeles, and Philadelphia (Bacon & Gomez, 2017; Manjoo, 2017). On the protest day, in Seattle-Tacoma Airport alone, about 3,000 protesters gathered to protest Trump's plan to ban citizens of certain Muslim countries from entering the US (Associated Press, 2017). In addition to Seattle, thousands of protesters also gathered in major airports in cities such as Portland (Quimby, 2017), Los Angeles (Edwards, 2017), and Philadelphia (Horn, 2017).

Activists called for people to gather with picket signs at various airports to draw media attention to their objections to Trump's immigration ban. These protest tactics were both more traditional and less disruptive than the tactics employed by "Day Without Immigrants." Differences in social media responses to these tactics may provide important practical implications for organizing future protests.

4.3 DATA COLLECTION

This dissertation aims to understand how protest is related to the change in online user's prejudice speech against immigrants. To achieve this goal, panel data were collected from 102, 094 users in Twitter. Panel data, or user historical posts, are embedded in each user's timeline profile, which displays what the user posted in the past. For each user, I collected all available tweets during the study period (two weeks before, two weeks after for each protest event). The following section provides a background description of the Twitter platform (see §4.3.1) and user panel data collection (see §4.3.2).

4.3.1 Twitter platform

A major challenge of studying the correlation between protests and prejudice is access to suitable data. Given that the occurrence of protests is not under the control of researchers, researchers in this area often rely on retrospective data collection. In the past, researchers have recruited protest participants and bystanders, and asked them to recall their experiences (Harlow & Guo, 2014). This data collection introduces recall bias where participants may not be able to accurately remember their experiences. Some research also has used secondary data sources where survey data were collected around protest events (Banaszak & Ondercin, 2016; Branton et al., 2015). While using secondary survey data may be promising, it requires researchers to rely on rare cases where surveys happened to be collected during the event period. To address these challenges, this study uses Twitter data as an observational field to examine prejudice change in the context of protests.

Twitter is a free social networking and micro-blogging platform, which enables users to post, send, and read each other's "tweets:" short messages composed of up to 140 characters. In total, Twitter has 119 million users worldwide, with 68 million users in the US (Rourke, 2017). Among the US users, Twitter is more popular among younger and more educated Americans; 36% of the users are ages 18 to 29, and 29% have college degrees (Greenwood et al., 2016). In addition, Twitter is popular among American minorities, with about one-in-four Latinos and Blacks using the site (Krogstad, 2015).

Twitter is also suitable for longitudinal text mining and analysis (Chew & Eysenbach, 2010), especially in the areas related to political issues or events (Lin, Margolin, Keegan, & Lazer, 2013). According to a recent nationally representative survey (Rainie, Smith, Schlozman, Brady, & Verba), 38% of Twitter users have promoted materials related to politics or social issues. This is more prevalent among users with political affiliations. A total of 52% liberal Democrats have used Twitter to promote materials supporting their political view; and 42 % conservative Republicans have also done so. In addition to promote political views, 31 % of users also have used Twitter to encourage other people to take action on a political or social issue that is important to them.

4.3.2 Panel data collection in Twitter

Panel data collection were carried out in March 2018. Users of interest were selected from multiple data sources and filtered based on inclusion criteria. For each selected user, all available tweets during the study period were collected from their timeline profile. This section describes data sources where users were selected from and inclusion criteria used for further filtering, followed by a description of how panel data was collected for each selected user (see Figure 3 for an overview of panel data collection).

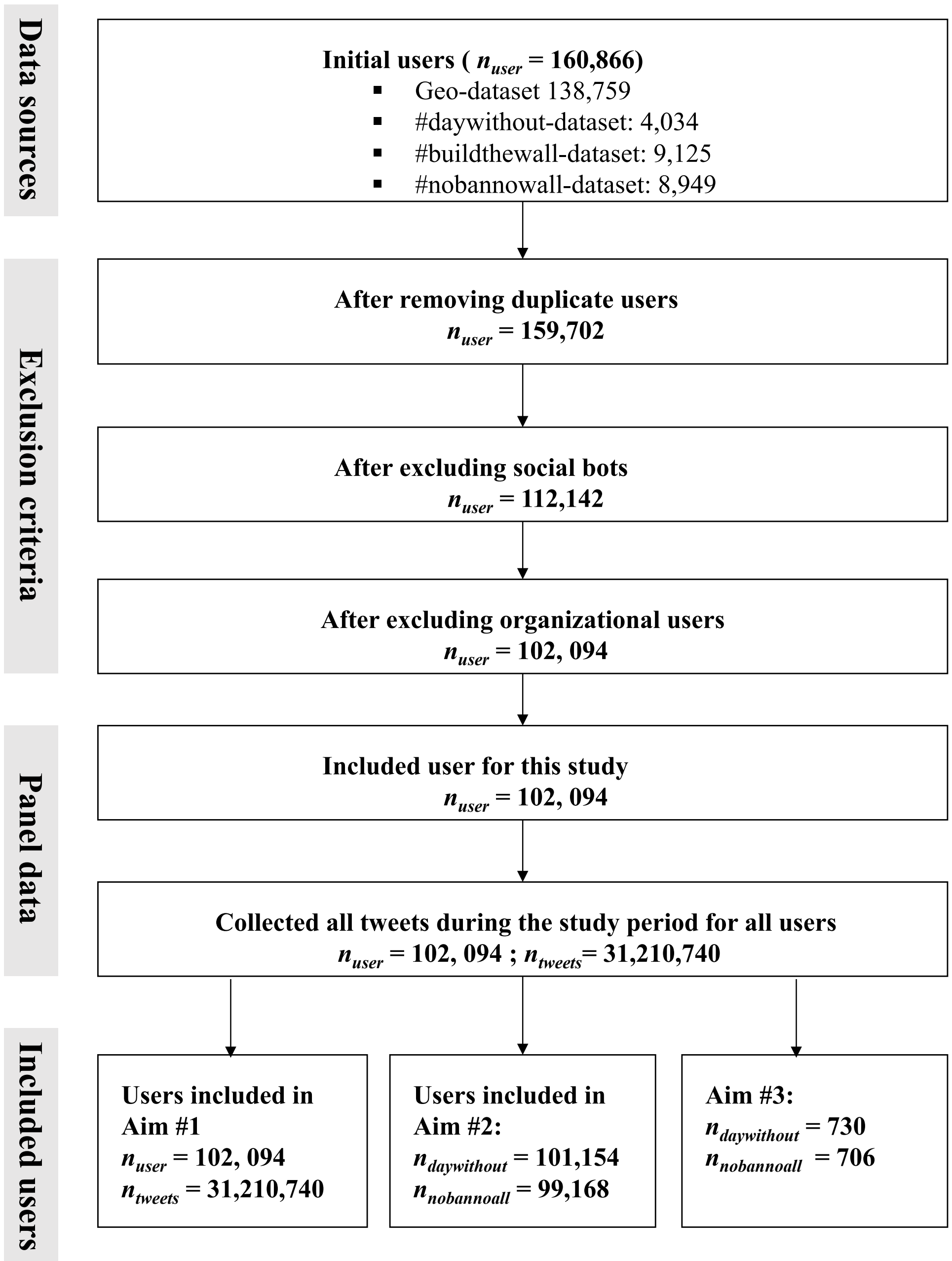

Figure 3 An overview of panel data collection

Data sources. Multiple data sources were used for selecting users of interest. These data were geo-based and hashtag-based datasets. Geo-based datasets were used as initial datasets because one of the study aims was to examine the relationship between levels of exposure to local protest and user's online prejudiced speech against immigrants. However, solely relying geo-tagged datasets posed risks to sampling bias. To mitigate the risks, I also collected additional data that contains protest event-related hashtags. The following section describes details about geo-based and hashtag-based datasets.

The geo-tagged dataset was provided by a research collaborator[3]. These geo-based users were included in the initial dataset because one of the study goal was to examine the role of geo-exposure in user's online prejudiced speech against immigrants. To achieve goal, one of the critical task is to exclude users who have never discussed immigrants in their tweets because users who never discussed immigrants in their tweets are unlikely to be the ones who explicitly express prejudiced speech against immigrants. To this end, I included users who were located in the US and showed interest in discussing topics that were relevant to immigrants. These topics were identified based on a set of keywords or keyword patterns related to immigrants: "latino," "mexican," "muslim," "islam,", or "immigra*," where "immigra*" match "immigrant," "immigrants," and "immigration." The keywords ("latino," "mexican," "muslim," "islam") were used because they are relevant to specific immigrant groups that have been the focus on immigration discussions. The words "immigrant," "immigrants," and "immigration" were used because they refer to immigrants in general and immigration issues. Previous research has used these keywords to study the discussions of immigrants on Twitter (Chung et al., 2016). In total, there were 138,759 users included in this study from the geo-tagged dataset.

3 https://github.com/bianjiang/tweetf0rm.

Admittedly, solely relying on geo-based data introduced sampling bias because not all Twitter users choose to disclose their geographic locations. To mitigate this risk, I collected additional users who showed interest in discussing topics related to the protest events. These users were identified based their mentions of #DayWithoutImmigrant," "#NoBanNoWall," and "BuildtheWall". Twitter API (a programming tool for collecting Twitter data) was used to users who mentioned these hashtags. In total, I identified 22,108 users. Among these users, 4,034 unique users mentioning "#DayWithoutImmigrant"; 8,949 unique users mentioning "#NoBanNoWall", and 9,125 unique users mentioning "#BuildtheWall".

Exclusion criteria. Exclusion criteria were used to remove duplicated users, social Bots, and organizational users. Since users were identified from multiple source, I first removed users who appeared more than once in the data. After removing these users, 159,702 users remained in the data.

Social Bots are accounts controlled by software that automatically generates contents (Varol, Ferrara, Davis, Menczer, & Flammini, 2017). Varol and colleagues (2017) estimated that between 9% and 15% of active Twitter accounts are bots. Given the interest of this study is human users, and thus bots were excluded prior to data analysis. To remove the social bots from these users. I used Botometer detection system API. The following section provides details about this system and procedures used to remove bots.

The Botometer detection system API[4] was used to obtain bot scores by evaluating the extent to which a Twitter account exhibits similarities to the characteristics of a social bot (Varol et al., 2017). It is a supervised machine learning algorithm that is trained on 1,150 features derived

4 botometer.iuni.iu.edu

from user profile, content, and social networks such as distribution of followers and friends. The system has been widely used in previous studies (Shao, Ciampaglia, Varol, Flammini, & Menczer, 2017). This model has high accuracy in differentiating human and bot accounts, with an area under the curve (AUC, a summary measure of the accuracy of a model prediction) of 94%. Varol et al. (2017) also showed that a classification score above 0.5 can be interpreted as a bot, which yields 86% overall classification accuracy. An classification model that has an overall accuracy above 80% is considered good (Habryn, 2014). Therefore, a threshold of 0.5 was used to classify bots and non-bot users. After excluding social bots, a total of 112,142 users remained.

In addition to remove bots, organizational user accounts were also removed from this study. A user is considered to have an organizational account if the account represents an institution, corporation, agency, news media, or common interest group (Oentaryo, Low, & Lim, 2015). To identify such accounts To identify such accounts, I used the machine learning tool developed by McCorriston, Jurgens, and Ruths (2015). To build the tool, the researchers created a dataset of manually classified accounts from a representative sample of Twitter and then used a Support Vector Machine, a supervised machine-learning classifier, to classify between organizational and personal accounts. This tool yielded an 88% overall classification accuracy. After excluding organizational accounts, a total of 102, 094 users remained.

Panel data collection. Panel data were collected for each included user. In Twitter, panel data is embedded in user's timeline profile, which displays the latest tweets from the specified (public) Twitter account. At the time of data collection, Twitter REST API (Application Programming Interface, a programming tool for collecting Twitter data) had a limitation for returning up to 3,200 tweets of a user's most recent Tweets[5]. For each of the user, Twitter REST

[5]https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html

API was used to collect *all available tweets* during the study period (two weeks before, two weeks after for each protest event). In total, I collected a total of 31,210,740 tweets posted during the study period from all included users (n_{users} = 102, 094).

Users included for each study aim. In this dissertation, study aim #1 focused on developing the measure for online prejudiced speech against immigrants. To achieve this aim, all of users and their tweets were used in the measurement development.

Study aim #2 focused on examining the role of temporal and geographic exposure to protest in online prejudiced speech against immigrants. To achieve this aim, users were excluded if geo-location information (see§ 4.4.1 for geographic exposure measure) was not identifiable or matched multiple geo-locations during the study period. In total, 101,154 users were included for the case study of the "Day Without Immigrants" protest, and 99,168 users were included for the "No Ban, No Wall" protest.

Study aim #3 focused on examining the role of group identity to protest in online prejudiced speech against immigrants. To achieve this aim, users were excluded if their group identity (see §4.4.2 for group identity measure) was not identifiable based on their user profiles. In total, 730 users were included for in the case study of the "Day Without Immigrants" protest; and 706 users were included for the "No Ban, No Wall" protest.

For both study aim #2 and aim #3, the sample sizes are much smaller than the collected panel data. This is especially true for aim #3. This is due to incomplete user's profiles in Twitter, where users did not provide geo-location or user's group identity in their profile. I further discussed this limitation in the discussion section.

4.4 MEASUREMENT

To understand the impact of protest on online prejudiced speech, this dissertation used different methods to measure exposure to protest events (see §4.4.1), and users' group identity (see §4.4.2), and online users' prejudiced speech against immigrants (see §4.4.3). The following section describes methods used for each measure.

4.4.1 Exposure to protest

4.4.1.1 Temporal Exposure to protest

Temporal exposure to protest was measured by a binary variable coded *1* if the users' tweets appeared within two-weeks *after the protest began* and coded *0* if the users' tweets appeared within two-weeks *before the protest began*.

4.4.1.2 Spatial exposure

Spatial exposure to protest was measure at three levels: high, medium, and low. The construction of high, medium, and low exposure to protest was to test hypothesis 2: Higher levels of exposure to local protest are associated with less prejudiced speech against immigrants among social media users. Specifically, users were defined as high exposure if they were located in the cities where the protests happened; as medium exposure if they were not located in any of the protest cities but were located in one of the states in which protests occurred; and as low exposure if they were located within the US but not in any of the cities or states where protests occurred.

To estimate spatial exposure to protest for each user, I need to know 1) user's geo-location during the study period; 2) a list of cities where protest took place. In Twitter, users have the option to enable location services on their account. Once this service is enabled, users can geo-tag their tweets with precise location in the form of the longitude and latitude. This location information was used to establish the geographic context in which users were located during the study period. This information was leveraged to measure user's geographic proximity to cities where local protest events took place (hereafter, protest cities).

Specifically, user's geo-locations were obtained using a Python programming package named tweet_parser. One of the methods in this package is called tweet_parser_getter_methods.tweet_geo[6]. This method gets user's derived location data in the form of the longitude and latitude and categorizes it into discrete geo-location representation in the form of city, state. For example, given the longitude and latitude of a user's tweet is [40, -105], this method categorizes this into [Boulder, Colorado, US]. This method was used to identifying geo-location information for each user in study aim #2.

In addition to identifying user's geo-location information, a list of protest cities was obtained to construct spatial exposure to protest. Specifically, I used news reports about the "Day Without Immigrants" and "No Ban, No Wall" protest as a source to identify a list of cities were these protest events took place. Wikipedia, a free online encyclopedia, was first used to gather references to news reports about these events. Wikipedia was used a tool to because previous study has shown that the content of Wikipedia has high accuracy (87%), despite the fact that its contents were edited by members of the general public, rather than experts (Chesney, 2006). For each identified news resource, I search their original reference, read the new reports, and identified protest cities reported by the news. The "Day Without Immigrants" took place in 17 cities from 11

6 See the description of "user.derived.locations.full_name" in the Twitter Geo 2.0 documentation: https://developer.twitter.com/en/docs/tweets/enrichments/overview/profile-geo.

states; the "No Ban, No Wall" protest took place in 22 cities from 15 states. Cities in which both protests occurred are the following: Atlanta (Georgia), Austin (Texas), Dallas (Texas), Boston (Massachusetts), New York (New York), Philadelphia (Pennsylvania), San Francisco (California), and Washington DC. Details about identified protest cities and their referenced news reports are included in Appendix A.

4.4.2 Group identity

Group identity was measured based on user's self-description in their user profile. A user was coded as immigrant if she or he explicitly mentioned their immigrant identity, such as "a first-gen immigrant, college-grad". A user was coded as non-immigrants if she or he was explicit about their American identity, such as "American by birth, Tennessee Christian by the grace of God". User's immigrant identity and non-immigrant identity were used because the interest of this study in in-group and out-group dynamics. According to intergroup emotion theory, compared to non-immigrant users, immigrant users might be less likely to express prejudiced speech against immigrants after the protests.

The user's group identity was obtained from user's profile. In Twitter, users have the option to provide information about themselves such as their race, gender, and occupation. This information is recorded in the user description field[7]. This field was used in this study because it provides straightforward information about user's identity and has minimal risk for including information that users are not fully disclosed to the public.

7 See the description in the Twitter documentation: https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object.

For each user's description, I focused on description that contain immigrant-related keywords (procedure for identifying the keywords and all included keywords see 4.4.3.2 and Table 4.2). For each description that contain the keywords, I read their description and labeled the users as immigrants if they explicitly mentioned their immigrant identity in the profile.

To construct non-immigrant users, I used keywords "American", "Americans", and "US born" to identify users who explicitly mentioned that they were born in the US. These keywords were used because they were the most straightforward keywords to identify users who self-identified as non-immigrant. For each description that contained these words, I read their description and labeled the users as non-immigrants if they explicitly mentioned that they were Americans or they were born in the US.

4.4.3 Online user's prejudiced speech against immigrants

The goal of this dissertation is to understand the impact of protest on online user's prejudiced speech against immigrants. To achieve this goal, one of major challenges is to reliably measure online user's prejudiced speech against immigrants. To address this challenge, I designed a two-step process (See Figure 4):

- *Step 1: identify prejudiced tweets.* For each tweet, I leveraged machine learning techniques to classify whether a tweet is about immigrants and whether a tweet is prejudiced speech against immigrants.
- *Step 2: estimate user-level prejudiced speech.* Specifically, I measured the intensity of prejudiced speech against immigrants at user level on a bi-weekly basis (two weeks before and two weeks after a protest event). For a user i at given time t, the user's prejudiced speech against immigrants is defined as the percentage of total number of tweets that were

prejudiced speech against immigrants among the total number of tweets that were about immigrants[8]. In this dissertation, t is defined in a two-week time period before and after a given protest event. The reason for using percentage rather than raw count is to account for the change in user's interest in discussing immigrants related to the protest.

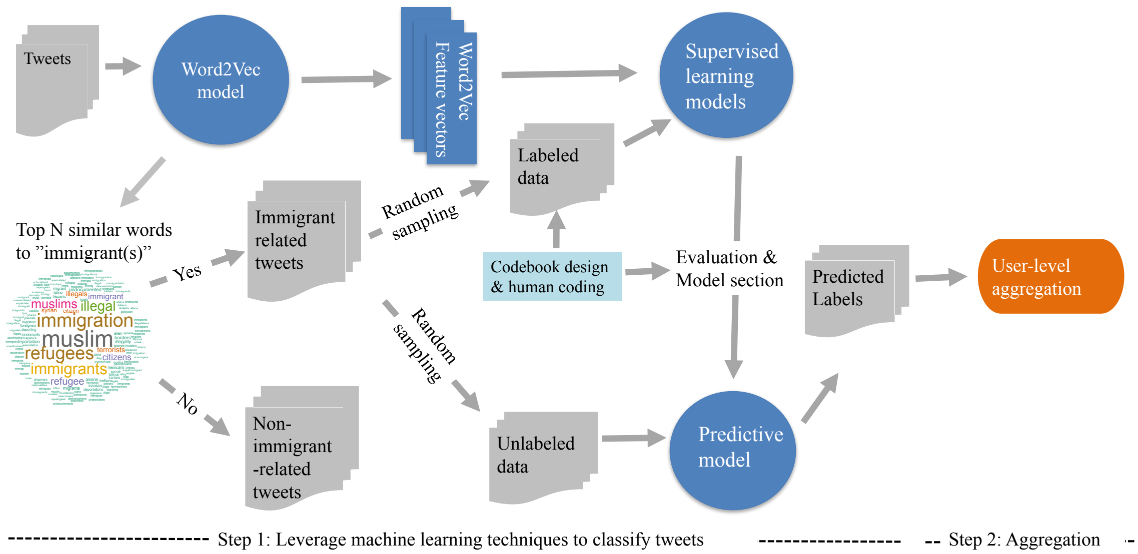

Figure 4 Two-step process for measuring online user's prejudiced speech against immigrants

Here is an overview of methods that I used in the step 1 in this process. The goal of step 1 is to classify whether a tweet is about immigrants and whether a tweet is prejudiced speech against immigrants. This requires classifying all collected tweets (n_{tweets} = 31,210,740) posted during the study period from all included users (n_{users} = 102, 094). To achieve this goal, Word2Vec was first used to train vector representations of words so that words with similar meanings are placed in closer numeric proximity. The purpose of this step is to identify words that share similar meanings with "immigrants" whereby the tweets relevant to immigrants are extract based on these keywords. After identifying relevant tweets, I hired human coders to label a random sample. This purpose of

[8] To preserve the distribution and avoid divide by 0 error, ε (10e-10) was added to the denominator.

labelling tweets is to evaluate a variety of supervised learning algorithms and select the one that performed the best. Last, this best performed algorithm was then used to automatically label tweets.

In following section, I first provide a background description of machine learning and Word2Vec models used in this work (see §4.4.3.1). Then, I describe how I used Word2Vec to identifying immigrant-related tweets (see §4.4.3.2). Last, I discuss how supervised learning models were used to automatically label tweets (see §4.4.3.3).

4.4.3.1 Machine learning techniques used in this study

Machine learning techniques use algorithms or statistical methods to classify, predict, and discover hidden patterns in a large quantity of data. Past research has shown that machine learning can automate human labeling work in an efficiently and reliably (Grimmer, 2015). I applied two types of machine learning techniques: word embedding model developed by Mikolov, Chen, Corrado, and Dean (2013) and supervised learning techniques. The key idea is to use a machine learning model to identify a collection of tweets that are relevant to immigrants, from which supervised learning techniques are trained on labeled data and used to automatically classify whether a tweet is about immigrants and whether a tweet is prejudiced speech against immigrants. The following section provides an overview of these techniques and their specific usage in this study.

Word2Vec, is a neural-network-based machine learning model used to convert semantic descriptions of words to numeric vectors by examining natural language and learning how words are used in a particular body of text (Mikolov et al., 2013). It creates vector representations of words so that words with similar meanings are placed in closer numeric proximity. For example, a Word2Vec model shows that the similar words *shocked*, *appalled* and *astonished* are numerically closer because these words often are used in a similar context. In this study, the Word2Vec model was used to expand immigrant-related keywords (see §4.4.3.2) and to incorporate them into featured vectors to train supervised learning models (see §4.4.3.3)

Supervised learning is a type of machine learning approach, consisting of first constructing ground truth (e.g., tweets labeled by human coders), and then applying an algorithm to the ground truth to predict the correct response when posed with new information (e.g., unlabeled tweets). In this study, supervised machine learning models were used to classify whether a given tweet was about immigrants, and whether it was prejudiced speech against immigrants. Since supervised learning requires pre-labeled ground truth, I first developed coding instructions to guide human coders for the classification tasks. These coding instructions was then tested for its reliability and used to label a random sample of 3000 tweets (see §4.4.3.4). To identify appropriate algorithms, I split the labeled tweets (ground truth) into 60% training, 20% development, and 20% test set, and conducted a series of experiments to test the performance of multiple supervised learning models to classify whether a tweet was about immigrants, and whether it conveyed prejudices, respectively. The model accuracy was evaluated based on F1-score, precision, recall, and AUC. AUC measures how well or accurate a model predicts in a classification task. Recall relates to false negative rate, and high recall correlates to a low false negative rate. Precision relates to a false positive rate, and high precision correlates to a low false positive rate. F1-score is the weighted average of precision and recall, measuring how much compromise between precision and recall is worthwhile. The models that performed the best were used to classify the remaining data.

4.4.3.2 Word2Vec and immigrant-related keywords expansions

Identifying relevant information to focus on is the key to effectively classify all collected tweets (n_{tweets} = 31,210,740) posted during the study period from all included users (n_{users} = 102, 094). This is because it is impossible and not cost efficient to manually label each tweet, and then aggregate tweets to user level. To be cost efficient, I leveraged a keyword query approach that can quickly determine tweets that are relevant to immigrants. The following section describe the details about this approach.

Keyword query is an approach to identify relevant information. Appropriate keywords can help filter noise and is relatively cost-effective. For example, by searching "immigrants" in Twitter, we can obtain a collection of tweets that are relevant to immigrants. One of the challenges in keyword query is to identify appropriate keywords. Narrowly defined keywords might result in the problem of missing relevant information, whereas broadly defined keywords might result in too much noise. To address this challenge, I adopted the idea of keyword expansion used in previous research (Olteanu, Castillo, Boy, & Varshney, 2018), a set of initial keywords were used to bootstrap words similar to these words. In this study, I used "immigrant" or "immigrants" as initial keywords and applied Word2Vec model to identify words that are similar to these words, followed by manual inspection. The following provides details for how I conducted keyword expansion to identify immigrant-related keywords and extract immigrant-related tweets.

Specifically, "immigrant" or "immigrants" were used as initial keywords. To expand the keywords, I first trained a Word2Vec model using the Gensim[9] package to construct the semantic description of words as numeric vectors. The model was trained on pre-processed data where stop words and URLS were removed. The trained model provided a 300-dimension vector for each word. Then, the top most similar words to the initial query were identified based on their ranked similarity score. The score was obtained by computing the cosine similarity between the vectors of initial queries and the vectors of words in the trained model. Higher cosine similarity score

9 https://pypi.org/project/gensim/

suggests greater similarity. To select the optimal, I used the top 20 as an initial state and incremented each iteration by 10. The iterative process ended when no words were identified at the top by manual inspection.

Inclusion criteria were created for manual inspection and to select the relevant keyword. This manual inspection step was taken because it might prove useful in future evaluations of the relevance of keywords based on previous research on topics related to immigrants. Specifically, in the manual inspection criteria, a word was included if it was a misspelled root word, synonym, word related to immigrant groups (e.g., Muslim and Mexican), or action relevant to immigrants (e. g, deportation and assimilation). In addition, I included figurative language referring to immigrants, i.e., metonymy, synecdoche, and metaphorical words. Figurative language was included in the manual inspection criteria because previous studies on discourse research show that discussions of immigrants have used figurative language such as metonymy to refer to immigrant groups. Metonymy replaces the name of a referent with the name of an entity which is closely associated (Blackledge, 2005; De Cillia, Reisigl, & Wodak, 1999). News media, for example, has constructed "illegal immigrant" as a metonym for Latino immigrants (Stewart, Pitts, & Osborne, 2011). Synecdoche, similar to metonymy, replaces the name of a referent with the name of another referent which is either semantically wider or semantically narrower (Blackledge, 2005; De Cillia et al., 1999). For example, people with antipathy toward immigrants have tried to employ synecdoche by replacing the description of immigrant as hardworking family with a description of the immigrant as criminal or terrorist (Stewart, 2012). Metaphor can be understood as the "transfer," "projection," or "mapping" from one experiential domain to another (Schäffner, 2002, p. 28). Metaphorically, immigration has been represented as an "invasion," "flood," and "burden" to the American society, something to be "weeded out" (Santa Ana, 1999; Wodak & Reisigl, 2001).

Table 1 Top n words to "immigrant" and "immigrants"

Initial words	immigrants			immigrant		
Top n	words included in the current step	similarity score range	Inclusion rate (%)	words included in the current step	similarity score range	Inclusion Rate (%)
20	immagrants, refugees, illegals, aliens, muslims, migrants, criminals, citizens, immigrates, emigrants, mexicans, legals, imigrants, immigra, immigran, dreamers, alliens, immigrantion,	0.79-0.63	90	imigrant, immagrant, undocumented, undocu, refugee, muslim, immigrants, unassimilated, immig, expat, imm, immigr, immigration, alien, migrant, immigrantion, immigran	0.74-0.57	85
30	imm, somalis, immigr, invaders, terrorists, undocumented, immigrant, refug	0.63-0.6	87	undoc, emigration, immigrats,iligal,inmigrant	0.57-0.54	73
40	imms,undoc,immigrats,immigrantscall,illegal,immigration,deportation,foreigners	0.59-0.58	85	illigal, hispanic, immi, immagrants	0.52-0.50	65
50	immig, syrians, residen, refugee, latinos, expelling	0.57-0.54	80	deportation	0.50-0.50	54
60	muslim, deportations, illegalaliens, illigal, farmworkers, rapefugees	0.54-0.53	77	schooljihadi, asylee, illegal	0.50-0.50	50
70	inmigrants, imigrant, hispanics, freeloaders, repatriation, invasions, citzens	0.52-0.52	76	iranian,immigant	0.50-0.50	46
80	laborers,cubans	0.52-0.51	69	haitian	0.49-0.49	41
90	moslems, imagrants, latinxs, rapists, imagration	0.51-0.50	67	rapefugee,alliens,refugees	0.48-0.49	40
100	jihadists, unassimilated, arabs, borders, emigration, terroists	0.50-0.49	66	sibrian,afghani,unau,indian,latinx	0.48-0.48	41
110	immigrations,iraqis	0.49-0.49	62	migrates	0.47-0.47	38

120	ilegal, palestinians, lawbreakers, musli, assimilation, assimilating	0.49-0.48	62	illeg,pakistani,invader	0.47-0.47	38
130	immgrnts, undesirables, nationalities, descendants, undocumenteds, immigrati	0.48-0.47	62	refug,dreamer	0.47-0.46	36
140	immigs, rapeugees, muslins	0.47-0.47	59	citizen,dacamented,influx	0.46-0.46	36
150	iranians,illegally	0.47-0.47	57	syrian, immigrantscall, latino, honduran, llegal	0.46-0.46	37
160	imigration, refuges	0.47-0.47	54	ethiopian	0.46-0.46	35
170	immigrts, pakistanis, deporting, rapefugee, settlers	0.47-0.46	54	uzbek	0.45-0.45	34
180	migrant, newcomers, migration, dacamented	0.46-0.46	53	imms, somali, somalis	0.45-0.45	33
190	muslin, nigerians, jihadis	0.46-0.46	52	expatriate, immigrantwho	0.44-0.44	33
200	undocu, filipinos	0.45-0.45	51			
210	latinx, kenyans, ctzns, deportees, murderers	0.45-0.44	50			
220	salvadorans, israelis, illegalls, immigrat	0.44-0.44	50			
230	llegal, assimilated	0.44-0.44	49			
240	infiltrators, afghans, extremists, lprs	0.44-0.43	48			
250	italians, wetbacks, undocume, scroungers	0.43-0.43	48			
260	roundups, illigals, illeg, biafrans, immegrants	0.43-0.43	48			
270	expats	0.43-0.43	47			
280	immagration, unauthorised, illega	0.43-0.43	46			
290	illgls	0.42-0.42	45			

Table 1 shows the words most similar to “immigrant” and “immigrants.” The keywords identified at each step are sorted in descending order based on their similarity score to the initial words. The similarity score range shows maximum and minimum scores. The inclusion rate is calculated based on the percentage of included words after manual inspection among all top words.

Using the keyword expansion approach, I identified a total of 161 unique immigrant-related keywords. These keywords were then used to extract immigrant-related tweets. In total, there were 889,579 immigrant-related tweets from 71,919 Twitter users. For immigrant-related tweets, I further classified them as tweets about immigrants, and tweets prejudiced against immigrants.

4.4.3.3 Supervised learning and automatically label tweets

To classify whether a given tweet was about immigrants, and whether it was prejudiced speech against immigrants I used the supervised learning approach, which consists of first constructing ground truth where tweets were manually labeled by human coders, followed by evaluating a variety algorithms to select the one that performed best. This algorithm was then used to automatically label tweets. The following section details the procedures for constructing ground truth, evaluating algorithms, and classifying tweets.

4.4.3.4 Constructing ground truth

The key to constructing ground truth is to develop a reliable codebook. It is a set of instructions that guide human coders for labeling tweets. The following provides details for codebook design, coding process, and codebook evaluation.

Codebook. In this study, a codebook was developed for human coders to classify whether a tweet was about immigrants, and whether it contained prejudiced speech against immigrants. The codebook included a definition of immigrant, derived from Zong and Batalova (2015), and examples and rationales for coding a tweet to be about immigrants. The indicators and examples were derived from coding a random sample of tweets. Rationales were brief descriptions of reasons for coding a sample tweet as being about immigrants. The same structure was used in the coding instruction for classifying whether a tweet was prejudiced against immigrants.

The definition and indicators of prejudiced speech against immigrants were derived from previous research (Allport, 1954; Klöckner & de Raaf, 2013). Prejudiced speech against immigrants was operationalized as a tweet conveying antipathy or hostility toward immigrants. For example, a prejudiced tweet that insults immigrants could be expressed in the following ways, "All You Fuckin Immigrants Fucked," or "We must not 'normalize' Mexican Invaders, we must call them out. Do not use 'immigrant' avoid it, deport all of them". Both tweets express prejudice against immigrants because they used words (e.g., fuckin and invaders) to curse and stigmatize immigrants or a specific immigrant group. In addition to insults, there are other indicators such as criticizing immigrants for damaging the economy, culture, or public safety (e.g., "Check out these jerk immigrants stealing your jobs."), and opposing immigration by supporting punitive immigration laws and/or distorted views about immigrants (e.g., "#DayWithoutImmigrants is great, DECADES without immigrants would be even better, they should all go back to their home") Appendix C provides details about the codebook used to code prejudiced speech.

Coding process. Two independent coders were recruited to assist the coding process, both being native English speakers of college-level education who are active social media users and check social media posts every day. Two training sessions were conducted before coders were asked to code the tweets independently.

In the first training session, I provided a brief overview of the study, discussing coding tasks and overall work flow. Following the training session, coders were asked to code a random sample of 200 immigrant-related tweets that had already been coded by the author based on the codebook; this batch was used to facilitate training. In the second training session, I discussed the coding results with the coders, and we each explained our reasons for the answer codes in the batch. Through the discussion, I found that misclassifications were due primarily to the misinterpretation of the tweets. For example, one coder misclassified this tweet, “FBI's pre-election sweep of Muslim Americans raises surveillance fears,” to be about immigrants. The tweet is about Muslim Americans, not Muslim immigrants. After the discussion, we reviewed our coding. The final codes for this batch were based on majority rule, and this batch was then used as the gold standard for future coding. After training sessions, coders proceeded to code four batches of 200 randomly sampled tweets. Each batch was coded independently by the coders. The results of these batches were used to test codebook reliability.

Evaluation. Cohen’s kappa coefficient (k), a statistic that measures interrater reliability for qualitative items, was used to evaluate coding results. It expresses a value between 0 and 1, with a high score indicating higher reliability. Landis and Koch (1977) defined the following interpretation system which can work as a general rule of thumb:

- < 0: Less than chance agreement
- 0.01–0.20: Slight agreement
- 0.21– 0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–0.99: Almost perfect agreement

Coding results. Table 2 shows the reliability scores for coding tweets about immigrants, and those that contain prejudiced speech against immigrants. Both coders had substantial agreement ($k > 0.61$) with the author on the gold standard batch before they proceeded to test codebook reliability, which consisted of four batches (Test 1 - 4). Coders had substantial agreement in coding each batch, which indicates that the codebook achieved high reliability.

Table 2 Interrater reliability for coding results

		Cohen's kappa	
Batch	coders	About immigrant	Prejudiced speech against immigrants
Gold standard	coder 1 vs. author	0.95	0.82
	coder 2 vs. author	0.83	0.75
Test 1	coder 1 vs. coder 2	0.72	0.63
Test 2	coder 1 vs. coder 2	0.86	0.88
Test 3	coder 1 vs. coder 2	0.87	0.61
Test 4	coder 1 vs. coder 2	0.93	0.77

Note. The sample size for each batch is 200.

Following the codebook reliability testing, coders and the author coded an additional 2000 tweets. The average pairwise Cohen's kappa coefficient was 0.87 for coding tweets about immigrants, and 0.63 for coding tweets that exhibited prejudiced speech against immigrants. The majority rule was used to decide the final code for each tweet. In total, there were 3000 labeled tweets. Table 3 shows the distribution of labeled tweets. A total of 1717 tweets (about 50%) were labeled as about immigrants. A total of 471 (about 16%) were prejudiced against immigrants.

Table 3 The distribution of labeled tweets

	The number of tweets labeled as:	
	Yes	No
About immigrants	1717	1283
Prejudiced speech against immigrants	471	2529

4.4.3.5 Experiment setup

Experiments were carried out to select machine learning models for classifying tweets about immigrants, and tweets that contained prejudiced speech against immigrants. Both were binary classification tasks with the objective of classifying whether a tweet belonged to one category or the other. Prior to training classification models, text pre-processing was performed on both labeled and unlabeled tweets to remove noise and prepare the text for classification. In this process, I removed stop words, URLS, and mentions (@username). The labeled tweets were split into 60% as training, 20% as test, and 20% as development, a common practice in machine learning.

The mean vectors of the Word2Vec model were used as features (or predictors) to train the classification models. Specifically, each tweet T consists of words $w_1, w_2, \ldots, w_n$. After training the Word2Vec model, each word is represented in a 300-dimensional vector, $u_{w1,} u_{w2,\ldots,} u_{w300.}$ The mean vectorization of the embedding model for a given tweet is defined as taking the average of all the word vectors in the tweet.

Supervised machine learning models. In the experiments, I tested the following supervised machine learning models: Naive Bayes, Adaptive Boosting, Support Vector Machines, Logistics Regression, and Extreme Gradient Boosting. These models were chosen because they have been shown to perform well in classifying tweets (Davidson, Warmsley, Macy, & Weber, 2017; He et al., 2017). Naive Bayes is a probabilistic classifier based on applying Bayes' theorem, which describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Adaptive Boosting (AdaBoost) is an ensemble classifier composed of multiple classifier algorithms (Freund & Schapire, 1997); it retains the algorithm by selecting the training dataset based on the performance of previous training. Support Vector Machines (SVM) is a discriminative classifier defined by a separating hyperplane a plane divided into two parts inn twodimensional space. Logistic Regression, borrowed from the field of statistics, estimates probabilities of the labels that one intends to predict, which is then transformed into discrete binary values for classification. Extreme Gradient Boosting (XGBoost) is a tree boosting classifier designed and optimized for boosted tree algorithms (Chen & Guestrin, 2016). The implementation of these algorithms was carried out using Sciki-learn and XGBoost Python packages.

Imbalanced data and over-sampling. As shown in Table 4.3, there was a major issue with imbalanced data where only about 16% were labeled prejudiced against immigrants. Previous research has shown that classification of data with imbalanced class distribution can suffer significant drawbacks in model performance because most standard classifier learning algorithms assume a relatively balanced class distribution and equal misclassification costs. This would lead classifiers to be more sensitive to detecting the majority class and less sensitive to the minority class (Sun, Wong, & Kamel, 2009). To address the issue of imbalanced tweets that contained prejudiced speech against immigrants, I used the naive random over-sampling technique to generate tweets that were labeled as prejudiced speech (the minority class). This over-sampling technique generates new samples by randomly sampling the replacements of the current available samples. The over-sampling was only applied to the training dataset. The random over-sampling was implemented using the imbalanced-learn Python package.

4.4.3.6 Evaluation metrics

Accuracy of the models was determined based on precision, recall, F1-score, and AUC. Precision measurement is defined as the number of true positives over the number of true positives plus the number of false positives. High precision therefore correlates to a low false positive rate. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives; high recall correlates to a low false negative rate. F1-score is the weighted average of precision and recall, measuring how much compromise between precision and recall is worthwhile, reaching its best value at 1 and worst at 0. Precision is defined as good if it is above 80%; recall is considered good if it is above 70%; F1-score is considered good if it is above 75% (Habryn, 2014). AUC, the area under the ROC (receiver operating characteristic) curve, is a summary measure of the accuracy of a model prediction. The AUC of 0.91 to 1 represents an excellent prediction; 0.81 to 0.9 represents a good prediction; 0.71 to 0.8 represents a fair prediction; 0.6 to 0.7 represents a poor prediction, and below 0.6 shows the overall model prediction was poor (Tape, 2001).

4.5 ANALYSIS STRATEGY

The unit of analysis in this dissertation focuses on user-level (individual-level). In study aim #1, the best performed (based on accuracy measure) model was selected to automatically label whether a tweet is about immigrants; and whether a tweet is prejudiced speech against immigrants. Then, I aggregated tweets to user-level. For each user, I calculated the percentage of tweets that were prejudiced against immigrants among the total number of tweets that were about immigrants. Study aim #2 and aim #3 applied this measure (developed in aim #1) to examine 1) the role of temporal and spatial exposure to protest in user's prejudiced speech against immigrants; and 2) the role of group identity in the relationship between protest and user's prejudiced speech against immigrants.

Compute before and after differences. To compare group differences, I computed the differences in the average intensity of prejudiced speech against immigrants for each user (p_j) in a pre-protest (14-day) and post-protest (14-day) window for each individual, $\Delta p_j = \frac{\sum_{t_e \leq t < t_{28}} p_j^{(w)}(t) - \sum_{t_0 \leq t < t_e} p_j^{(w)}(t)}{m}$, where t_e is the time of the protest event.

Significance test. The paired samples Wilcoxon test (also known as Wilcoxon signed-rank test) was then used to test the significant level in the differences. This test was chosen due to the non-normal distribution of user's prejudiced speech against immigrants. The paired samples Kruskal-Wallis test was conducted to further determine whether the changes in user's prejudiced speech against immigrants from baseline before the protest began to after the protest began were associated to different levels of exposure to local protests. Table 4 provides an overview of data analysis for each aim. The major difference between the Wilcoxon Signed-Rank and the Kruskal-Wallis test is that the latter can accommodate more than two groups.

Table 4 Data analysis and aims

Aim 1: Develop a method for measuring online prejudiced speech against immigrants.	
• Machine learning models were built to classify whether a tweet is about immigrants; and whether a tweet is prejudiced speech against immigrants. • For each user, I calculated the percentage of tweets that were prejudiced against immigrants among the total number of tweets that were about immigrants.	
Aim 2: Examine the role of temporal and spatial exposure to protest in online prejudiced speech against immigrants	
<u>Hypothesis 1:</u> Social media users express less prejudiced speech against immigrants after protests compared to baseline before protests.	• Compute before and after differences • The paired samples Wilcoxon test
<u>Hypothesis 2:</u> Higher levels of exposure to local protest are associated with less prejudiced speech against immigrants among social media users.	• Compute before and after differences • The paired samples Kruskal-Wallis test
Aim 3: Examine the role of group identity in the relationship between protest and online prejudiced speech against immigrants.	
Hypothesis 3: After the protest, social media users who were self-identified as immigrants express less prejudiced speech against	• Compute before and after differences • The paired samples Wilcoxon Test • The paired samples Kruskal-Wallis test

immigrants compared to users who were self-identified as non-immigrants.	

5.0 RESULTS

The goal of this dissertation is to understand the impact of protest on online prejudiced speech against immigrants. Specifically, I leverage machine learning models for measuring online prejudiced speech (Study aim #1). Then, I applied this measure to explore the role of temporal and geographic exposure to protest in online user's prejudiced speech (Study aim #2). Last, I applied this measure to explore the role of group identity in the relationship between protest and online user's prejudiced speech (Study aim #3).

In §5.1, I describe the accuracy for applying machine learning model in classifying whether a tweet is about immigrants; and whether a tweet is prejudiced speech against immigrants. In §5.2, I describe the results for the relationship between temporal and geographic exposure to protest in online user prejudiced speech against immigrants. In §5.2, I describe the results for the relationship between protest and online prejudiced speech among user groups.

5.1 AIM #1: MEASUREMENT ACCURACY FOR ONLINE PREJUDICED SPEECH

In aim #1, I leveraged machine learning models to measure user's online prejudiced speech against immigrants. Accuracy of the models were assessed at tweet-level. Specifically, for each select models, I evaluated their accuracy with respect to classifying whether a tweet is about immigrants; and whether a tweet is prejudiced speech against immigrants.

Table 5 shows the accuracy of supervised learning models for classifying the following categories: 1) whether a tweet is about immigrants; and 2) whether a tweet is prejudiced speech against immigrants. For category 1), both AdaBoost and SVM had good precision (above 80%), recall (above 80%), and F1-score (above 80%), suggesting that the selected features in combination with the models were able to retrieve most of the tweets that were about immigrants and had few false positives. In addition, both models also reached an AUC above 0.8, showing that they were reliable prediction models for classifying whether a tweet was about immigrants. When comparing AdaBoost with SVM, the overall performance of SVM was slightly better, with 1.6% performance gain over AdaBoost for F1-score. Therefore, the performance of SVM was the best among all evaluated models.

For category 2), the overall performance of AdaBoost and XGBoost was better than the other models. Both of these models reached good precision (above 80%), recall (above 80%), and F1-score (above 80%), and AUC (above 0.8), suggesting that these models were able to reliably classify whether a tweet was prejudiced speech against immigrants. When comparing XGBoost with AdaBoost, the overall performance of XGBoost was slightly better, with a 1.5% of performance gain over AdaBoost for F1-score. Therefore, the XGBoost performed the best among all evaluated models.

Table 5 Classification performance of supervised learning models across categories

Category	F1-score (%)	Precision (%)	Recall (%)	AUC	Method
Whether a tweet is about immigrant	82.9	83.0	82.9	0.824	AdaBoost
	73.8	74.5	73.5	0.736	Naive Bayes
	74.1	78.3	76.0	0.712	XGBoost
	84.5	**84.9**	**84.4**	**0.846**	**SVM**
	75.2	78.1	76.7	0.724	Logistic Regression
Whether a tweet is prejudiced speech against immigrant	83.4	84.3	82.7	0.722	AdaBoost
	74.9	83.3	71.3	0.714	Naive Bayes
	84.9	**85.4**	**84.5**	**0.737**	**XGBoost**
	79.8	85.1	77.3	0.754	SVM
	66.6	85.1	61.5	0.720	Logistic Regression

After selecting the best performing models, experiments were carried out to understand how much labeled data was sufficient to obtain a reasonable result. Therefore, I experimented with the system by incrementally adding batches of instances, such as 200, 600, 1000, 1400, 2200, and all instances. This method could help future researchers employ the model at the onset of a protest event when sufficient amount of labeled data is available (Alam, Joty, & Imran, 2018).

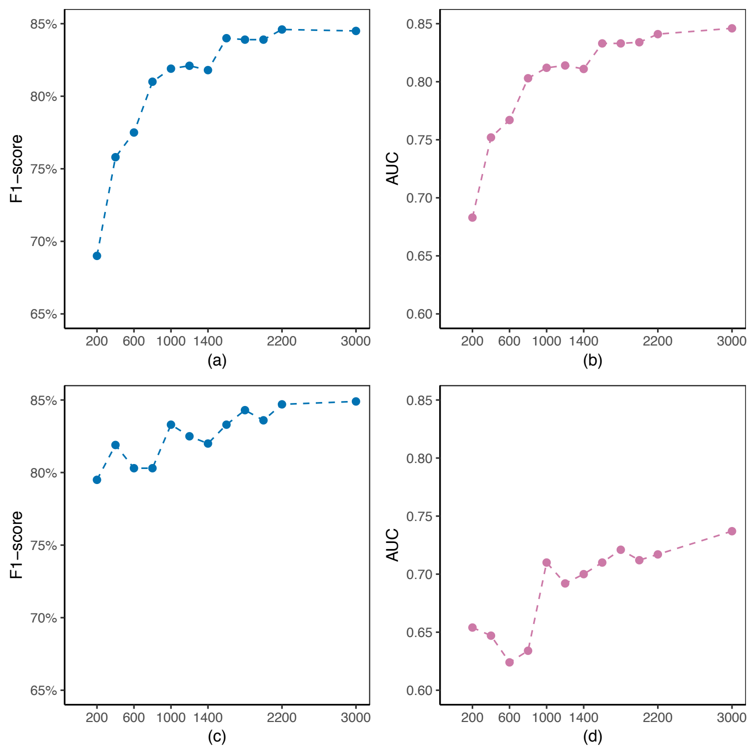

Figure 5 F1-score and AUC for the selected supervised models using different batch sizes

Figure 5 (a) and (b) show the changes for classifying tweets that are about immigrants; Figure 5 (c) and (d) show the changes for classifying tweets that are prejudiced speech against immigrants.

Figure 5 displays changes of F1-score and AUC using different batch sizes across different classification categories. Figure 5 (a) and (b) show the changes for classifying tweets about immigrants using SVM; and Figure 5 (c) and (d) show the changes for classifying tweets that are prejudiced speech against immigrants using XGBoost. For the task of classifying tweets about immigrants, the model's performance improved as more labeled data were included— from 69% to 85% for F1-score and from 0.68 to 0.85 for AUC. Specifically, major improvements were observed when enlarging thebatch size from 200 to 1400, beyond which, however, the performance improvements were comparatively minor. The results obtained using 1000 and 1400 were in the acceptable range where F1-score was above 80% and AUC was greater than 0.8. For the task of classifying tweets that are prejudiced speech against immigrants, the model's performance improved as more labeled data were included— from 80% to 85% for F1-score and from 0.65 to 0.73 for AUC. In this case, major improvements were observed when batch size was enlarged from 200 to 2200. Despite the performance trend showing that more labeled data could potentially increase the model performance, the results obtained using 1400 and 2200 were in the acceptable range where F1-score was above 80% and AUC was greater than 0.7. Therefore, it was observed that at least 1400 labeled data was needed to obtain reasonable results for the classification tasks.

To automatically label the unlabeled data, I used the trained SVM to label whether a tweet is about immigrants and the trained XGBoost to label whether a tweet is prejudiced speech against immigrants. Among 889,579 immigrant-related tweets, 490,622 tweets (about 55%) were labeled as about immigrants, and 157,014 (about 18%) were labeled as prejudiced speech against immigrants.

5.2 STUDY AIM #2: EXPOSURE TO PROTEST AND ONLINE PREJUDICED SPEECH

In this study, I explored the role of temporal and geographic exposure to protest in online prejudiced speech using two protest events: the "Day Without Immigrants" protest and the "No Ban, No Wall" protest. For each event, I conducted non-parametric test to analyze the differences in Twitter user's intensity of prejudiced speech before and after each protest event two protest events and its intensity across users located in protest cities, protest states, and outside of protest states. Non-parametric test was chosen because the intensity of prejudiced speech was not normally distributed. Below provide details about analyses and results for each protest event.

5.2.1 Case study 1: "Day Without Immigrants"

The "Day Without Immigrants" protest took place on February 16, 2018. On the protest day, activists called for immigrants regardless of their legal status to stay home from work or school, close their businesses, and abstain from shopping. To understand the differences in online user's prejudiced speech against immigrants before and after the protest, users were selected from the study sample based on the time of their tweets, and they were included if they posted tweets two weeks before and two weeks after the "Day Without Immigrants" protest began on February 16, 2017. A total of 101,154 users were included in the analysis.

Temporal exposure. Table 6 shows the mean and standard deviations for the total number of tweets in the protest event window, the total number of tweets about immigrants, the total number of tweets that showed prejudiced speech against immigrants, percentage of tweets about immigrants, and online users' prejudiced speech against immigrants.

On average, the average number of tweets decreased after the protest (*M* =71) compared to the baseline before the protest (*M* =74). While there was an increase in the number of tweets about immigrants after the protest (*M* =1.33) compared with the baseline before the protest (*M* =1.29), the percentage of tweets about immigrants among total tweets remained the same. In addition, compared with the baseline before the protest (*M* = 9%), there was an 1% increase in user's intensity of prejudiced speech against immigrants after the protest (*M*= 10%).

Table 6 Distribution of tweets in protest event window at user-level

	"Day Without Immigrants" protest (n=101,154)		
	Mean	SD	Median
The total number of Tweets:			
before	74	101	42
after	71	101	40
The number of tweets about immigrants per user:			
before	1.29	4.49	0
after	1.33	4.53	0
The number of tweets prejudiced against immigrants per user:			
before	0.48	2.34	0
after	0.54	2.50	0
The percentage of tweets about immigrants per user:			
	2%	0.05	0.0%
	2%	0.05	0.0%
The intensity of prejudiced speech against immigrants per user:			
before	9%	0.27	0.0%
after	10%	0.28	0.0%

Wilcoxon Signed-Rank test was used to test whether there was a significant difference before and after protests began in terms of online user's intensity of prejudiced speech against immigrants; this test was chosen due to the non-normal distribution of users' intensity of prejudiced speech against immigrants. The test showed that although users' prejudiced speech against immigrants was significantly more intense after the protest began than before the protest began, z = -11.86, p <0.001, r = 0.04, the effect size[10] (r = 0.04) was insubstantial according to Cohen (1988).

10 The effect size r is calculated as based on equation 2.18 in (Rosenthal, 1991).

In sum, while there was a difference between the baseline before and after protests began in terms of online user's prejudiced speech against immigrants, the magnitude of this difference is insubstantial.

Spatial exposure. Users with identifiable geo-locations were selected and partitioned into three groups (high, medium, and low exposure group) based on their geographic proximity to the cities where protests happened during the "Day Without Immigrants" protest. In total, 98,875 users were included in the analysis.

Table 7 shows the mean and standard deviation for the total number of tweets in the protest event window, the number of tweets about immigrants, the number of tweets that contained prejudiced speech against immigrants, and the intensity of prejudiced speech across low, medium, and high exposure groups. Overall, there were fewer tweets after the protest began compared with the baseline before the protest began across low ($M_{before} = 77.29$, $M_{after} = 75.23$), medium ($M_{before} = 60.68$, $M_{after} = 58.51$), and high exposure groups ($M_{before} = 58.54$, $M_{after} = 56.95$). However, this pattern was not observed for the number of tweets about immigrants, the number of tweets that contained prejudiced speech against immigrants, and the intensity of prejudiced speech against immigrants across low, medium, and high exposure groups.

After the protest began, the average number of tweets about immigrants was higher than the baseline before protest across low, medium, and high exposure groups. On average, the low exposure group posted more tweets about immigrants (M = 1.46) compared to the medium (M= 0.63) and high exposure groups (M = 0.67). Similarly, the number of tweets that contained prejudiced speech against immigrants was higher among low exposure group (M = 0.60) compared to the medium (M = 0.22) and high exposure groups (M = 0.22). In addition, the low exposure group (M_{after} =11.1%) also expressed higher intensity of prejudiced speech against immigrants compared to the high exposure group (M_{after}=7.9%).

Table 7 Distribution of tweets by levels of exposure to protest

	Low exposure (*n* = 86,716)			Medium exposure (*n* = 9,930)			High exposure (*n* = 2,229)		
	Mean	SD	Medium	Mean	SD	Medium	Mean	SD	Medium
The total number of tweets in window:									
before	77.29	103.75	44	60.68	75.53	36	58.54	78.45	35
after	75.23	104.00	43	58.51	75.43	34	56.95	76.01	32
The number of tweets about immigrants per user:									
before	1.41	4.69	0	0.61	2.19	0	0.72	1.98	0
after	1.46	4.82	0	0.63	2.05	0	0.77	2.22	0
The number of tweets prejudiced against immigrants per user:									
before	0.53	2.48	0	0.20	1.09	0	0.18	0.89	0
after	0.60	2.67	0	0.22	1.00	0	0.22	0.86	0
The percentage of tweets about immigrants per user:									
before	1.6%	0.05	0	1.1%	0.05	0	1.5%	0.05	0.0%
after	1.8%	0.05	0	1.5%	0.06	0	2.0%	0.07	0.0%
The intensity of prejudiced speech against immigrants per user:									
before	10.0%	0.27	0.0%	6.0%	0.23	0.0%	6.0%	0.25	0.0%
after	11.0%	0.29	0.0%	7.0%	0.25	0.0%	8.0%	0.25	0.0%

The Wilcoxon test was used to test whether there was a significant difference across low, medium, and high exposure groups after protests began by controlling baseline before protest. For the low exposure group, it was found that users' prejudiced speech against immigrants was significantly more intense after the protest began than before the protest began (z = -12.56, p <0.001), though the effect size for this difference was insubstantial (r =0.04). A similar pattern was also found for the medium exposure group: while users' prejudiced speech against immigrants was significantly more intense after the protest began than before the protest began (z = -2.97, $p <$ 0.01), the effect size was trivial (r =0.03). For the high exposure group, users' prejudiced speech against immigrants was significantly more intense after the protest began than before the protest began (z = -2.97, p < 0.01). The effect size for this difference was small (r =0.1).

The Kruskal-Wallis test was conducted to further determine whether the changes in user's intensity of prejudiced speech against immigrants from baseline before the protest began to after the protest began were associated to different levels of exposure to local protests. The results of the test found significant differences across all the groups, $H(2)$ = 6.75, p < 0.05. Pairwise comparisons using Bonferroni correction showed that a statistical significance existed between low- and medium-exposure group (p <0.01). Effect size was calculated using the method in Tomczak and Tomczak (2014) and results suggest the differences were trivial: 0.00005. These findings showed that while changes in user's intensity of prejudiced speech against immigrants from baseline before the protest began and to after the protest began were related to different levels of exposure to local protests different level of exposure, the strength of this relationship is weak.

5.2.2 Case study 2: "No Ban, No Wall"

The "No Ban, No Wall" protest took place on January 28, 2017 as a response to President Donald Trump's plan to ban citizens of certain Muslim countries from entering the US, and suspend admission of all refugees entering the country. To understand how users expressed prejudice against immigrants differently before and after the protest, I included users who posted tweets in the two-week periods before and after the "No Ban, No Wall" protest took place, including a total of 99,168 users in the analysis.

Temporal exposure. Table 8 shows the mean and standard deviations for the total number of tweets in the protest event window, the total number of tweets about immigrants, the total number of tweets that showed prejudiced speech against immigrants, percentage of tweets about immigrants, and intensity of prejudiced speech against immigrants.

Table 8 Distribution of tweets in protest event window at user-level

	"No Ban, No Wall" protest (n = 99,168)		
	Mean	SD	Medium
The total number of tweets			
before	70.01	91.96	41
after	80.14	107.42	46
The number of tweets about immigrants per user			
before	0.99	2.98	0
after	2.25	5.86	0
The number of tweets prejudiced against immigrants per user			
before	0.27	1.23	0
after	0.58	2.39	0
The percentage of tweets about immigrants per user:			
	1%	0.04	0.0%
	3%	0.06	0.0%
The intensity of prejudiced speech against immigrants per user			
before	8%	0.26	0.0%
after	10%	0.27	0.0%

The average number of tweets increased after the protest began (M =80.14) compared to the baseline before the protest (M =70.01). There was a 2% increase in the average user's discussion of immigrants and a 2% increase in their intensity of prejudiced speech against immigrants after the protest compared to the baseline before the protest.

Wilcoxon Signed-Rank was used to test whether there was a significant difference before and after protests in the intensity of prejudiced speech against immigrants. The test results showed that users' prejudiced speech against immigrants was significantly more intense after the protest than before the protest, $z = -25.73$, $p < 0.001$. The effect size for the difference was small, $r = 0.1$.

Spatial exposure. Users with identifiable geo-locations were selected and partitioned into three groups (high, medium, and low exposure group) based on their geographic proximity to the cities where protests occurred. A total of 97,690 users were included in this analysis.

Table 9 Distribution of tweets in protest event window at user-level by exposure-level

	Low exposure (n = 81,959)			Medium exposure (n = 11,579)			High exposure (n =4,152)		
	Mean	SD	Medium	Mean	SD	Medium	Mean	SD	Medium
The total number of tweets in window									
before	72.79	94.98	43	61.13	75.06	37	59.09	73.38	35
after	84.64	112.10	49	64.60	80.61	39	63.11	76.26	37
The number of tweets about immigrants									
before	1.08	3.18	0	0.57	1.73	0	0.65	1.82	0
after	2.48	6.27	0	1.24	3.19	0	1.29	3.07	0
The number of tweets prejudiced against immigrants									
before	0.29	1.29	0	0.18	0.88	0	0.13	0.72	0
after	0.64	2.56	0	0.34	1.36	0	0.22	0.92	0
The percentage of tweets about immigrants									
before	1%	0.04	0	1%	0.04	0	1%	0.04	0.0%
after	3%	0.06	0	2%	0.07	0	3%	0.08	0.0%
The intensity of prejudiced speech against immigrants									
before	8%	0.26	0.0%	7%	0.26	0.0%	6%	0.23	0.0%
after	11%	0.27	0.0%	9%	0.27	0.0%	7%	0.22	0.0%

Table 9 shows the distribution of tweets in the protest event window of users by exposure-level. Overall, there were more tweets after the protest began compared with the baseline before the protest began across low (M_{before} = 72.79, M_{after} = 84.64), medium (M_{before} = 61.13, M_{after} = 64.60), and high exposure groups (M_{before} = 59.09, M_{after} = 63.11). Similar patterns were observed for the number of tweets about immigrants, the number of tweets that contained prejudiced speech against immigrants, the percentage tweets about immigrants, and the intensity of prejudiced speech against immigrants across low, medium, and high exposure groups.

After the protest began, the average number of tweets about immigrants was higher than the baseline before protest across low, medium, and high exposure groups. On average, the low exposure group posted more tweets about immigrants (M = 2.48) compared to the medium (M= 1.24) and high exposure groups (M = 1.29). Similarly, the number of tweets that contained prejudiced speech against immigrants was higher among the low exposure group (M = 0.64) compared to the medium (M = 0.34) and high exposure groups (M = 0.22). In addition, while the average percentage of tweets about immigrants was the same between low exposure and high exposure groups after the protest, the low exposure group (M =11%) expressed a higher intensity of prejudiced speech against immigrants compared to the high exposure group (M =7%).

Wilcoxon Signed-Rank was used to test whether there was a significant difference before and after protests within all the groups. For the low exposure group, it was found that users' prejudiced speech against immigrants was significantly more intense after the protest than before the protest (z = -25.20, p <0.001). The effect size for this difference was small (r =0.1). A similar pattern was also found for the medium exposure group, though while users' prejudiced speech against immigrants was significantly more intense after the protest than before the protest (z = -7.34, p < 0.01), the effect size was small (r =0.1). For the high exposure group, users' prejudiced speech against immigrants was significantly more intense after the protest than before the protest (z = -3.05, p < 0.01). The effect size for this difference was trivial (r =0.05).

The Kruskal-Wallis test was used to determine if the changes in user's intensity of prejudiced speech against immigrants from baseline before and after the protest were associated with different levels of exposure to local protests. The results of the test found significant differences across low, medium, and high exposure groups in changes in the intensity of prejudiced speech against immigrants, $H(2) = 24.24$, $p < 0.001$, 0.002. Pairwise comparisons using Bonferroni correction showed that a statistical significance existed between low- and medium-exposure group (p <0.001), and low- and high- exposure group (p <0.001). These findings showed that while the changes in user's intensity of prejudiced speech was related to different level of exposure, the strength of this relationship is weak.

5.3 STUDY AIM #3: GROUP IDENTITY, PROTEST, AND ONLINE PREJUDICED SPEECH

In this study, I explored the role of group identity in relationship between protest and online prejudiced speech using two protest events: the "Day Without Immigrants" protest and the "No Ban, No Wall" protest. For each event, I conducted non-parametric test to analyze the differences between immigrant and non-immigrant group Below provide details about analyses and results for each protest event.

5.3.1 Case study 1: "Day Without Immigrants"

A total of 730 Twitter users were included in this analysis, among whom, 321 users were self-identified as immigrants, and 409 users were self-identified as US-born Americans.

Table 10 shows the distribution of tweets about immigrants and prejudice against immigrants across these groups. On average, immigrants posted fewer tweets after the protest (M = 80.25) compared to baseline before the protest (M = 88.46). In contrast to immigrant users, US-born American users posted more tweets after the protest (M = 107.11) compared to baseline before the protest (M = 96.16).

For both immigrant and US-born American users, discussion about immigrants and the intensity of prejudiced speech against immigrants increased after the protest. Specifically, among immigrant users, the average discussion of immigrants increased 0.8% after the protest began (M = 5.3%) compared to baseline before the protest (M = 4.5%); and intensity of prejudiced speech against immigrants increased 2.2 % after the protest began (M = 19.1%) compared to baseline before the protest (M = 16.9%). Among US-born American users, there was a 0.4% increase in the average discussion of immigrants after the protest began (M = 3.4%) compared to baseline before the protest (M = 3.0%); and there was a 5.5% increase in the average intensity of expressing prejudiced speech against immigrants after the protest began (M = 28.2%) compared to baseline before the protest (M = 22.7%).

Table 10 User-level distribution of tweets about and prejudiced against immigrants

	Foreign-born immigrant group (n = 321)			US-born American group (n = 409)		
	Mean	SD	Medium	Mean	SD	Medium
The total number of tweets in window						
before	88.46	119.99	53	96.16	123.97	48
after	80.25	112.5	50	107.1 1	190.4	42
The number of tweets about immigrants						
before	4.07	8.21	1	3.12	6.08	1
after	3.91	7.6	1	3.68	10.47	1
The number of tweets prejudiced against immigrants						
before	1.39	4.77	0	1.61	3.71	0
after	1.39	3.76	0	2.09	6.43	0
The percentage of tweets about immigrants						
before	4.5%	0.073	0.8%	3.0%	0.069	2.2%
after	5.3%	0.095	1.1%	3.4%	0.065	1.9%
The intensity of prejudiced speech against immigrants						
before	16.9%	0.746	0.0%	22.7%	0.37	0.0%
after	19.1%	0.447	0.0%	28.2%	0.44	0.0%

Wilcoxon Signed-Rank test was also used to test whether changes in user's prejudiced speech against immigrants were related to users' group identity. For both groups of users, there was no significant differences in user's discussion of immigrants between baseline before the protest and after the protest. Within immigrant users, no significant differences were found in intensity of prejudiced speech against immigrants between baseline before the protest and after protest, z = -1.69, p = 0.09. However, within American users, user's intensity of expressing prejudiced speech against immigrants after protest was significantly higher compared with baseline before protest, z = -2.67, p < 0.01. The effect size for this difference was small ($r = 0.13$).

The Kruskal-Wallis test was conducted to further determine the significance of these differences between immigrant and American users in terms of their intensity of prejudiced speech against immigrants. Before the protest, no significant differences were found between immigrant and American users, $H(1) = 1.93$, $p = 0.16$. However, after the protest, American users expressed significantly higher intensity of prejudiced speech against immigrants than immigrant users, $H(1) = 1.93$, $p = 0.05$. Effect size (η^2) was calculated using method in Tomczak and Tomczak (2014) to estimate the magnitude of difference in immigrant and American users' prejudiced speech against immigrants after protests. The effect size was 0.005, suggesting that group identity explained 0.5 % of variance in users' prejudiced speech against immigrants after protests began.

5.3.2 Case study 2: "No ban no wall"

A total of 706 Twitter users were included in the analysis. Among these users, 318 users were self-identified as immigrants, and 388 users were self-identified US-born Americans.

Table 11 The distribution of tweets in protest event window at user-level by group identity

	Foreign-born immigrant group (n = 318)			US-born American group (n = 388)		
	Mean	SD	Medium	Mean	SD	Medium
The total number of tweets in window						
before	76.92	91.84	48	83.39	105.43	45
after	101.80	129.85	63	113.23	143.32	62
The number of tweets about immigrants						
before	3.15	5.72	0	1.88	3.81	1
after	6.16	10.21	2	4.91	8.28	2
The number of tweets prejudiced against immigrants						
before	0.59	1.36	0	0.72	2.15	0
after	1.48	4.04	0	1.96	3.99	0
The percentage of tweets about immigrants						
before	4%	0.07	0.0%	2%	0.03	1.7%
after	6%	0.07	2.5%	4%	0.08	4.3%
The intensity of prejudiced speech against immigrants						
before	13%	0.29	0.0%	19%	0.41	0.0%
after	15%	0.28	0.0%	25%	0.36	0.0%

Table 11 shows the distribution of tweets in the protest event window between the foreign-born immigrant group and the US-born American group. Both immigrant and US-born Americans posted more tweets after the protest. In addition, discussions about immigrants and the intensity of prejudiced speech against immigrants increased after the protest. Specifically, among immigrant users, the average discussion of immigrants increased 2% after the protest began (M = 6%) compared to baseline before the protest (M =4%); and the intensity of prejudiced speech against immigrants increased 2% after the protest (M = 13%) compared to baseline before the protest (M

= 15%). Among US-born American users, there was a 2 % of increase in the average discussion of immigrants after the protest (*M* = 2%) compared to baseline before the protest (*M* = 4%); and a 6% of increase in the average intensity of prejudiced speech against immigrants after the protest (*M* = 25%) compared to baseline before the protest (*M* = 19%).

Wilcoxon Signed-Rank test was used to test if changes in users' interest in discussing immigrants and intensity of prejudiced speech against immigrants were related to users' group identity. Immigrant's interest in discussing immigrants after the protest was significantly higher compared to baseline before the protest, $z = -10.86$, $p < 0.001$. The effect size for the difference (r =0.4) showed that the magnitude of the difference was large. However, there was no significant increase in immigrant users' intensity of prejudice after the protest compared to baseline before the protest ($z = -1.73$, $p = 0.08$). In contrast, both American users' interest in discussing immigrants ($z = -10.86$, $p < 0.001$, r = 0.4) and their intensity of prejudiced speech against immigrants increased significantly after the protest began immigrants ($z = -3.78$, $p < 0.001$, r = 0.2).

The Kruskal-Wallis test was conducted to further determine the significance of these differences between immigrant and American users' intensity of prejudiced speech against immigrants. Before the protest, no significant differences were found between immigrant and American users, $H(1) = 0.40$, $p = 0.52$. After the protest, however, American users expressed a significantly higher intensity of prejudiced speech against immigrants than immigrant users, $H(1) = 10.4$, $p < 0.01$, $\eta^2 = 0.014$. The effect size of 0.014 suggested that group identity explained 1.4 % of variance in users' prejudiced speech against immigrants after protests began.

6.0 DISCUSSION

Persistent problems with prejudice conveyed in public speech have perpetuated and intensified immigrants' experiences of being excluded from full participation in society (Fangen, 2010; Mullen & Rice, 2003). Today, one in seven Americans use social media to connect with each other, share information; many also use social media to promote protest and collective action (Pew Research Center, 2018). As we are increasingly living in online society, prejudiced speech has also moved online in an overt and extreme form (Davidson et al., 2017), yet its adverse effects are as real as offline speech (Kowalski et al., 2014). Immigrants were found to be particularly susceptible to the adverse effects of online prejudice (Comas-Forgas et al., 2017). While prejudice-reduction research has suggested the important role of social protests in suppressing stigmatizing public statements (Rüsch et al., 2005) and improving marginalized group's position and treatment (Abrams, Vasiljevic, & Wardrop, 2012; Dixon, Durrheim, Kerr, & Thomae, 2013; Dixon, Levine, Reicher, & Durrheim, 2012; Haslam, 2012), research to date however has not examined the role of protest in online prejudiced speech. One of the challenges for studying the role of protest in online prejudiced speech is to obtain or develop a reliable measure of outcome. While prior work proposed methods for measuring online hate speech, these methods have ignored the targets of hate speech (ElSherief et al., 2018) and are consequently not applicable for measuring online prejudiced speech against immigrants.

This dissertation sought to develop a method for measuring online prejudiced speech against immigrants, and leverage that method into the understanding of the relationship between protest and online prejudiced speech. Specifically, three studies were carried out: In **study 1**, I proposed a method for measuring online prejudiced speech against immigrants. In **study 2**, I focused on the role of temporal exposure and spatial exposure in social media users' prejudiced speech against immigrants. In **study 3**, I focused on the role of group identity in the relationship between protest and social media users' prejudiced speech against immigrants. Below I provide an overview of the study results, as well as a discussion of the limitations and implications for research and social work practice.

6.1 SUMMARY OF FINDINGS

This dissertation advances knowledge about the role of protest in online prejudiced speech against immigrants in two main ways. First, this research provides a reliable measure for online prejudiced speech against immigrants. Second, the findings of this research show that there is a correlation between protest and an increase in online prejudiced speech against immigrants. These findings suggest that there is a significant risk of unintended consequences arising from social protests. A detailed discussion of the research findings follows below.

6.1.1 A reliable measurement for online prejudiced speech

To evaluate interventions with an aim to reduce prejudices expressed online, one of the biggest challenges is to measure the phenomenon reliably. In this dissertation, I leveraged machine learning models into the development of a measurement for online prejudiced speech against immigrants. The proposed method started with defining a collection of Twitter posts that were relevant to a target group (in this study, immigrants). This approach addressed the limitations in current hate speech detection research, where research has focused on the use of offensive keywords to identify hate speech relevant information (e.g., Olteanu et al., 2018), but due to the confusion of offensive language and hate speech (Davidson et al., 2017) it has proven inapplicable to the study of hate speech that targets a specific group. While ElSherief et al. (2018) proposed a method that differentiated hate speech that targeted individuals and groups, this work failed to address the problem of identifying hate speech about a specific group. The method I propose in this this research addresses the problem, and future research in hate speech detection might consider applying the proposed method to the understanding of hate speech by defining the group relevant discussions, followed by hate and non-hate in the discussion of the group.

This research provided a valid and reliable codebook for coding prejudiced speech against immigrants, adopting the definition of immigrant from Zong and Batalova (2015) and the definition of prejudiced from Allport (1954). To showcase the indicators of prejudiced speech against immigrants, this codebook incorporated coding themes from Klöckner and de Raaf (2013). Two independent coders conducted reliability testing of the codebook resulting in an average kappa above 0.8, suggesting good reliability. Further research may adapt the codebook to the study of online prejudiced speech against immigrants.

This research leveraged machine learning models into the classification of tweets that contain prejudiced speech against immigrants. Sensitive analysis showed that, using word2vec features, the support vector machine learning model achieved high precision and recall (about 85%) in classifying whether a tweet was about immigrants, and the XGboost model achieved high precision and recall (about 85%) in classifying whether a tweet was online prejudiced speech against immigrants. The good performance of these models in conducting text classification tasks is consistent with prior research (e.g., He et al., 2017).

the methods employed in this research to test theories related to prejudice. For example, group threat theory hypothesizes that immigrant group size is related to anti-immigrant prejudice, with larger growth in immigrant group size leading to higher levels of prejudice against that group (Pottie - Sherman & Wilkes, 2017). However, it remains unknown whether this hypothesis is still true for online prejudice, because findings of prejudice research seem to suggest that in the US online prejudice against Muslim immigrants seems to be more prevalent than Mexican immigrants (Awan, 2014; Chung et al., 2016). In addition to testing theories, future research may also consider testing other types of interventions such as education. For example, researchers can develop an online education program and use the method developed in this research to measure the impact of this program on reducing online users' prejudiced speech against immigrants. Beyond these research questions, researchers might also consider applying the approach introduced in this research to quantify other types of prejudices such as prejudices against the homosexual population, racial prejudices, and prejudices against people with disabilities.

6.1.2 The role of protest in online prejudiced speech

This research conducted two case studies (the "Day Without Immigrants" and "No Ban, No Wall" protest) to examine the role of protest exposure and group identity in online prejudiced speech. No evidence was found for the role of protest in reducing online prejudiced speech against immigrants. Detailed discussions of the findings of each hypothesis follow below.

Hypothesis 1 in this study posed that social media users expressed less prejudiced speech against immigrants after protests compared to baseline before protests. The findings of this study did not support this hypothesis. Moreover, while the theory of civil resistance and long-run attitude change posed that nonviolent protest could shift attitudes in a positive way (Mazumder, 2018), this study did not find support for this premise. Results in both case studies showed that there was a significant increase in social media user's prejudiced speech against immigrants following the protest.

The increase in online prejudiced speech after the protest might also be explained by prejudiced norms. Research has shown that prejudice against immigrants is or has become socially acceptable in American society (Fisher et al., 2011; Nier et al., 2012; Zárate & Quezada, 2012). While protest may make the community at large more aware of the contenders, it may likewise revoke the need for the application of norms and values that might have been dormant or forgotten prior to protest (Coser, 1957). When prejudice is the norm in American society, it is possible that protest can intensify prejudiced attitudes against immigrants.

The prejudiced norm might be more prevalent in social media. While social media has been used to publicize local actions to distant audiences (Segerberg & Bennett, 2011), the combination of user anonymity, attention seeking, and other poorly understood dynamics of online social behavior might result in inflamed passions and viral propagation of blatantly prejudiced speech against immigrants even from users who were directly affected by local actions. As suggested in recent research, social media lacks the social restrictions and inhibitions that prevent people from speaking their prejudices in public (Lapidot-Lefler & Barak, 2012; Suler, 2004), and as a consequence, it creates a public space where people think it is socially acceptable to openly express prejudices of any kind (Spata, 2015). For example, evidence has shown that negative comments in social media received more attention than neutral comments (Brady, Wills, Jost, Tucker, & Van Bavel, 2017; Fan, Zhao, Chen, & Xu, 2014). Therefore, it is important for future researchers to consider social media dynamics in the study of using protest as a prejudice-reduction intervention. The findings of this research could be potentially used by activists, organizers, and protesters to design strategies and ways to mitigate risks associated with protest and to influence social media discussions of immigrants immediately after protest in a positive way.

Hypothesis 2 in this study posed that higher levels of exposure to local protest are associated with less prejudiced speech against immigrants among social media users. Overall, this research did not find evidence supporting this hypothesis. This may be due to the weight geography was given in this study. This research focused on city-level, because it provides a proxy for examining the relationship between distance to protest sites and users' responses. For example, on the "Day Without Immigrants" protest day, local businesses such as restaurants were closed in cities such as Philadelphia PA. It is possible that residents of Philadelphia would be affected more than residents who live in cities where protest did not happen such as Aliquippa PA and Charleston WV. However, it is possible that even among residents of the cities where protest happened, some might be affected more than others. As found by previous research, during the 2006 immigrant protests, Latinos located within 100 miles reported stronger feelings about their agency in bringing about political and social change (Wallace et al., 2014). Therefore, future research might need to examine the relationship between protest exposure and online prejudiced speech with a more fine-grained measurement of geography.

This study also found mixed results for the "Day Without Immigrants" and "No Ban, No Wall" protests in terms of the online prejudiced speech changes. After the "Day Without Immigrants" protest, the average users' intensity of prejudiced speech in the high exposure group was higher than that of users in the low exposure group. However, the "No Ban, No Wall" protest suggested contrary results: users' intensity of prejudiced speech in the high exposure group was significantly lower than that of users in the low exposure group. It should be noted that the differences in the tactics used by the "Day Without Immigrants" protest and "No Ban, No Wall" protest might also have played a role in how online users responded after the protest. The "Day Without Immigrants" protest primarily used boycott. As Wang and Piazza (2016) suggested, while the disruptive tactics of protest were effective in bringing public attention to issues of immigrants in a local context, these tactics might carry unintended consequences arising from social protests such as a risk of alienating support. On the protest day of "Day Without Immigrants", activists called for immigrants regardless of their legal status to stay home from work or school, close their businesses, and abstain from shopping. The tactics used in this protest might backfire by offending potential supporters who value work ethics or disapprove of behaviors such as truancy. In contrast, tactics (e.g., public demonstration) used in the "No Ban, No Wall" seem to be less extreme and less likely to offend supporters. Further research might consider sampling the same user cohort to further understand how different types of protest tactics (boycotts vs. marches) affect social media users' responses after the protest. In addition, it might also be important for protest organizers to consider weighing the trade-off between the intent to achieve protest goals and the risks associated with the protest tactics.

Hypothesis 3 in this study posed that after the protest, social media users who were self-identified as immigrants express less prejudiced speech against immigrants compared to users who were self-identified as non-immigrants. Findings of the case studies supported this hypothesis. Specifically, it was found that while immigrant users showed no significant changes in intensity of prejudiced speech against immigrants, American users expressed significantly more prejudiced speech against immigrants after the protest. These findings provide support for intergroup emotion theory, which proposes that people appraise ongoing events and entities through an in-group lens and respond to them with corresponding in-group emotions when their social identities are salient (Smith & Mackie, 2015). In this study, American and immigrant users were identified by their explicit mentions of being US-born Americans or being immigrants during the study period. The differences in the changes of intensity of prejudiced speech among American users and immigrant users might be related to their response with the protest through the in-group lens.

In addition, it was found that the magnitude of the changes in online prejudiced speech for American users after the "No Ban, No Wall" protest was larger than that after the "Day Without Immigrants" protest. This could be attributed to the differences in protest dates as discussed previously. It might also be related to the differences in baseline before the protest among American users. The intensity of online prejudiced speech at baseline before the protest was 19% for the "No Ban, No Wall" protest and 23% for the "Day Without Immigrants" protest. Beginning with a higher baseline, it might be possible that the change is less pronounced. In future research, it might be important to identify the same user cohort across different protest events to further understand how each protest contributes to social media users' prejudiced speech against immigrants over time.

Overall, this research did not find evidence that protest reduces online prejudiced speech against immigrants. This could be explained by the limited cases examined by this research. It might be possible that reducing online prejudiced speech against immigrants requires a series of protests that lasts months or even years. As suggested by previous research, civil rights movements lasted almost 15 years, and its impact on racial prejudices was found to be positive and long-lasting (Mazumder, 2018). In this case, future research could test the long-term effect of protest on reducing online prejudiced speech against immigrants.

It also remains unclear whether findings in this research call into question the effectiveness of protest in reducing prejudiced speech or are peculiar to social media. Past research that claims the effectiveness of protest in reducing prejudice has not considered online environment. It might be possible that the same suppression effect that protest has on prejudice (through face-to-face contact) has diminished as online social interaction increases. As suggested by previous researchers (Lapidot-Lefler & Barak, 2012; Suler, 2004), computer-mediated communication (compared to face-to-face communication) lacks the social restrictions and inhibitions that prevent people from speaking their prejudices in public. As a consequence, this environment may have created a public space where people think it is socially acceptable to openly express prejudices (Spata, 2015). It might be possible that when protest occurs, online users would respond in a way that they consider normal based on their social networks. Future research is needed to understand how online social networks or norms moderate the impact of protest on prejudiced speech.

Last, it should be noted that in addition to protest, past research has proposed other types of interventions that are known to be effective in reducing prejudice, such as contact and education. Research might consider for testing these interventions in the future research. For example, contact is an approach that allows members of different groups to resolve conflicts by communicating and learning to appreciate different points of view. While this intervention has been shown to be effective, it remains unknown whether it can also reduce online prejudiced speech. Further research, therefore, might consider testing the effectiveness of the contact approach in reducing online prejudiced speech. One idea is to develop community outreach programs in the areas (perhaps at the neighborhood level) found to have a high intensity of online prejudiced speech, and to observe whether its intensity is reduced after the implementation of the program. Additionally, researchers could also consider testing the effectiveness of education in reducing online prejudiced speech. The basic assumption of education-based interventions is that prejudice is the result of the individual's distorted thinking about out-group members due to their exposure to biased information in the environment. Therefore, to reduce online prejudiced speech, researchers might consider implementing educational programs that are intended to change users' exposure to biased information. In summary, while this research did not find evidence for the effectiveness of protest in reducing online prejudice, future researchers could apply the approach proposed by this research to gather further evidence, or explore the effectiveness of interventions such as contact and education in reducing online prejudice.

6.2 LIMITATIONS

Although the results of this dissertation hold implications for understanding the role of protest in online prejudiced speech against immigrants, it is necessary to note a number of limitations that temper the conclusion drawn from this research.

This research selected users from keyword-related geo-tagged tweets and tweets that contained specific hashtags in Twitter. This convenient sampling approach introduced potential selection bias where the sample included in this study might not reflect the general Twitter population. Future research might consider drawing a representative sample of Twitter users to address this limitation. Moreover, given that this sample of online users was restricted to Twitter, the conclusion of this work must not extrapolate to other social media platforms. For future research, it might be important to compare differences in how online users express prejudiced speech against immigrants before and after protests across different social media platforms.

The other major limitation of study sample is the way social media users' group identity is constructed; it is based on solely the social media user's profile. The user selection method may potentially exclude users who did not reveal their identity in their user profile but also identify with immigrants or non-immigrant groups. Therefore, it should be noted that the findings the role of group identity played in the relationship between protest and online prejudiced speech against immigrant may not be generalized to social media users who did not report their identity in user profile. Future research might consider developing a more general approach that can make inferences about users' identity beyond their user profile.

This research had no control over a number of confounding variables because protest events occurred in a natural environment. The increase in online prejudiced speech might be related to these factors: user's political ideology, immigrant population size in a given city or state, Twitter users' tendency to express negative opinions about issues of any kind, attention seeking behaviors, and social influence from other users. For example, previous research has shown that people who self-identified as conservatives were more likely to express prejudice toward immigrants compared to those who self-identified as liberals (Caricati et al., 2017; Chambers et al., 2013; Kugler et al., 2014; Van de Vyver et al., 2016). In addition, the link between conservative identity and social dominance orientation has been found in conservative political rhetoric; for example, conservative rhetoric is often accompanied by an emphasis on dangerous and competitive worldviews (Lakoff, 1997; Lane, 1962). Therefore, future research might consider controlling for political ideology as well as other factors in examining the relationship between protest and online prejudiced speech. In addition, the design of this study is a quasi-experimental design; the lack of random assignment of users into protest exposure group and non-exposure group might also threaten the internal validity of this research.

Prejudiced speech, similar to hate speech, is a challenging phenomenon to define and not monolithic. The classification of online prejudiced speech against immigrants might reflect the subjective biases from recruited coders and the author. Future research might consider testing the reliability of the codebook among different coders using posts from different social media platforms. In addition, while the results of machine learning models showed high reliability for the classification tasks, it is important to recognize potential social biases that entered into the algorithms. It might be important for future research to identify these pre-existing biases underlying the implicit values of our society, and correct the algorithms accordingly.

In addition, Mazumder's (2018) theory of civil resistance and long-run attitude change lacks empirical evidence. This theory was built on a single case by studying the impact of civil rights protest on racial prejudice. While this study provided support for the impact of protest on reducing racial prejudice in the long run, it did not rule out the possibility that protest could invoke backlash in a short-term. The lack of empirical evidence seems to be a general weakness in the study of protest as prejudice-reduction intervention. As Corrigan et al. (2012) suggested, while anecdotal evidence indicates that protest can be an effective approach in reducing prejudice and stigma, only a few studies have tested its effectiveness. To address the general issue of lacking empirical evidence, it is important for future research to provide more evidence with respect to the role of protest in prejudice-reduction intervention.

6.3 IMPLICATIONS

Despite the previously noted research limitations, the results of this dissertation provide some important implications for social work practice. First, because this research did not find evidence for the effectiveness of protest in reducing online prejudiced speech, this suggests that social work practitioners may need to be cautious when using protest as an intervention strategy to reduce online prejudiced speech against immigrants. Second, despite not finding evidence for the effectiveness of protest in reducing online prejudiced speech, this research did find that protest may increase discussions of immigrants on social media. This finding suggests that it might be important for social workers to consider the trade-off between using protest as a prejudice-intervention strategy versus policy intervention. Third, as this research provided a reliable tool for measuring online prejudiced speech, social workers might also consider using this tool to monitor prejudiced speech in social media, and possibly seek interventions that could be directly implemented on these online platforms. These implications are discussed in greater detail below.

Protest has been proposed as one of the major interventions for reducing prejudice. In the past decade, there has been a tremendous increase in immigrant protests both in the US and worldwide (Tyler & Marciniak, 2013). These protests involved actors such as undocumented workers and concerned citizens, humanitarian organizations, and non-profit organizations. The underlying assumption that motivates these protesters could be their belief about using protest to as a way to address injustice and to resist prejudice and discrimination encountered by immigrants. It is the same assumptions and beliefs motivated that many people to get involved in immigrant protests. For example, many participants of the "Day Without Immigrants" may have believed that by staying home, skipping school, or closing their business would change Americans' prejudiced against immigrants from thinking of immigrants as making no contribution to American society, to thinking of them as economic contributors to society. Indeed, past research also suggested that while exposure to protest movements may not change attitudes overnight, it can lead to attitude change over time (Banaszak & Ondercin, 2016; Mazumder, 2018). This, then, is evidence that supports the claim that protest should be used as a strategy for resisting prejudice and discrimination against immigrants.

Surprisingly, this research did not find any evidence supporting for the effectiveness of protest in reducing online prejudiced speech. On the contrary, the results of this study suggest that protest may invoke backlash and spark more prejudiced speech against immigrants among social media users. These results imply that practitioners may need to be cautious when using protest as an intervention strategy to reduce online prejudiced speech. For this reason, there has been increased awareness of, and efforts in, reducing online prejudiced speech across non-profit organizations and research institutions such as The Nexus Fund, Berkman Klein Center, and Media Awareness Network. The Nexus Fund is a non-profit foundation that supports efforts to combat prejudiced and hate speech against ethnic, religious, racial minorities, women, and the LGBTQ groups. A recent report from the Nexus Fund suggested protest to be one of the effective interventions in reducing prejudiced and hate speech, and listed "No Ban, No Wall" as a successful protest (The Nexus fund, 2019). However, this claim was not supported in this research. Therefore, it might be important for practitioners to take other research methods into consideration before concluding that protest is effective in reducing online prejudiced speech.

While this research did not find evidence for the effectiveness of protest in reducing online prejudiced speech, it does not make any claims about the overall effectiveness of protest in promoting social change. This is especially important for understanding the role of protest in policy change. As suggested by social impact theory, the impact of protest on social change is manifested not only in cultural changes (e.g., changes in prejudiced attitudes), but also in policy changes. In the past, research has shown that protest could lead to policy change. For example, the 2006 immigration protests were found to have an impact on foreign-born Latinos' immigration policy preferences about amnesty (Branton et al., 2015). In addition, immigrant protest was also found to raise the awareness of immigration related issues among the general population (Carey Jr et al., 2014). This finding is consistent with results in this research: it was found that there was a significant increase in the percentage of tweets about immigrants after the protest among social media users. For example, the medium score of the percentage of tweets about immigrants increased from 0.0% to 2.5% among immigrants and from 1.7% to 4.3% among US-born Americans. This means the majority of immigrant US-born Americans users in this study sample were more actively engaged in discussions about immigrants. This finding is important because bringing public attention to bear on issues of the moment is a necessary step in promoting policy change. Therefore, it might be important for practitioners and protestors to be mindful of the types of social changes that a protest is intended to achieve, and to properly assess the trade-off between using protest as prejudice-interventions and policy intervention.

It may be important also for social work to make effective use of social media in reducing prejudiced speech against immigrants. As we are living in an increasingly online society, the role social media plays in prejudiced speech, and interventions that intend to address the problem should not be overlooked. This research designed a reliable tool for measuring online prejudiced speech, which social workers might also consider utilizing to monitor prejudiced speech in social media, and seek interventions that might be implemented directly on these online platforms. As suggested by Blaya (2019), monitoring and analyzing prejudiced speech through automated techniques to identify and programmatically classify prejudiced and hate speech--such as analyzing its nature, prevalence, and trends--, is an important step for further intervening in such behaviors in online space. Collecting such data could help practitioners identify behavior and attitude patterns that would be used to implement prejudiced and hate speech prevention policies and interventions. This is especially important because tech companies such as Facebook, Twitter, and Google have policies that respond to abusive content . For example, Google recently has updated its policies to prohibit a wide range of content that could be discriminatory. This new policy not only prohibits content that advocates against a specific group, but also prohibits derogatory content that has any "characteristic that is associated with systematic discrimination or marginalization" (Google, 2019). In addition to the policy, these platforms also provide tools for reporting such content. Social workers might be able to use the tool developed in this research to monitor prejudiced and hate speech, and directly report the content to these tech companies so the policies can be enforced. In addition, social workers could also provide related resources to inform immigrants of these existing policies, and coordinate this reporting process in their daily interactions with immigrants.

Additionally, social workers could also seek interventions such as public denunciations, which might be useful in creating a public shaming effect on people who make stigmatized and prejudiced comments about immigrants (Casados, 2017; Goldman, 2015; Wahl, 1997). For example, social workers might employ social media encouraging the use of individual protest through publicly declaring themselves as #IAmStigmafree. Meanwhile, social work can also cooperate with humanitarian and non-profit organizations to use social media as a storytelling channel, creating different stories tailored to different audiences. This strategy has been shown to be effective in reducing prejudice against LGBT groups (Jones, 2015). The same strategy could potentially reduce prejudiced speech against immigrants. Using social media as a storytelling channel makes it possible to target different groups at the same time. For example, social work can leverage YouTube videos and create different types of stories with intended effects in mind such as videos with a shaming effect or videos invoking commonality between different immigrant groups. These videos can be sent spontaneously to different groups to enable massive, rapid, and spontaneous public responses to the issues of prejudice and discrimination faced in immigrants' everyday lives.

6.4 CONCLUSION

This dissertation sought to explore the role of protest in online prejudiced speech against immigrants. One of the challenges for studying the role of protest in online prejudiced speech is to obtain reliable measurement of outcome. To address this challenge, this research leveraged machine learning techniques into the development of the measurement for online prejudiced speech against immigrants, and applied this method to understanding the relationship between protest exposure and group identity in online prejudiced speech against immigrants.

This research advances knowledge about the role of protest in online prejudiced speech against immigrants in two main ways. First, it provided a reliable measurement for online prejudiced speech against immigrants. Second, the findings of this research showed that protest was related to the increase in online prejudiced speech against immigrants, especially among users who did not identify themselves as immigrants. These findings suggested a significant risk of unintended consequences arising from social protests. It is hoped that these findings will lead to continued progress on the part of social work researchers and practitioners to incorporate methods of data science into the understanding of social media dynamics and to seek interventions that can ultimately reduce prejudices that prevail in online space.

BIBLIOGRAPHY

Aboud, F., & Miller, L. (2007). Promoting peer intervention in name-calling. *South African Journal of Psychology, 37*(4), 803-819.

Aboud, F., Tredoux, C., Tropp, L., Brown, C. S., Niens, U., & Noor, N. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review, 32*(4), 307-336.

Abrams, D., Vasiljevic, M., & Wardrop, H. M. (2012). Prejudice reduction, collective action, and then what? *Behavioral and Brain Sciences, 35*(06), 425-426.

Acar, Y. G., & Uluğ, Ö. M. (2016). Examining prejudice reduction through solidarity and togetherness experiences among Gezi Park activists in Turkey. *Journal of Social and Political Psychology, 4*(1), 166-179.

Ackerman, E. F. (2013). The Rise of the "Illegal Alien". *Contexts, 12*(3), 72-74. doi:10.1177/1536504213499883

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York City, NY: W.W. Norton Company.

Akrami, N., Ekehammar, B., & Bergh, R. (2011). Generalized prejudice: Common and specific components. *Psychological Science, 22*(1), 57-59.

Alam, F., Joty, S., & Imran, M. (2018). Domain Adaptation with Adversarial Training and Graph Embeddings. *arXiv preprint arXiv:1805.05151*.

Aldana, R. (2007). Of Katz and aliens: Privacy expectations and the immigration raids. *UC Davis L. Rev., 41*, 1081.

Allport, G. W. (1954). *The nature of prejudice*. Boston, MA: Addison-wesley.

Alsultany, E. (2012). *Arabs and Muslims in the Media: Race and Representation after 9/11*. New York City, NY: NYU Press.

Alsultany, E. (2013). Arabs and Muslims in the media after 9/11: Representational strategies for a "postrace" era. *American Quarterly, 65*(1), 161-169.

Altemeyer, B. (1981). *Right-wing authoritarianism*. Manitoba, Canad: University of Manitoba press.

Amenta, E., & Young, M. P. (1999). Making an impact: Conceptual and methodological implications of the collective goods criterion. *How social movements matter*, 22-41.

Andrews, K. T. (1997). The impacts of social movements on the political process: The civil rights movement and black electoral politics in Mississippi. *American sociological review*, 800-819.

Andrews, K. T., Beyerlein, K., & Tucker Farnum, T. (2015). The legitimacy of protest: explaining White Southerners' attitudes toward the civil rights movement. *Social Forces, 94*(3), 1021-1044.

Arredondo, A., Orozco, E., Wallace, S. P., & Rodríguez, M. (2012). Health insurance for undocumented immigrants: opportunities and barriers on the Mexican side of the US border. *The International journal of health planning and management, 27*(1), 50-62.

Asbrock, F., Christ, O., Duckitt, J., & Sibley, C. G. (2012). Differential effects of intergroup contact for authoritarians and social dominators: A dual process model perspective. *Personality and Social Psychology Bulletin, 38*(4), 477-490.

Associated Press. (2017). More Than 3,000 Protest Trump Executive Order At Sea-Tac Airport. *KIRO7* Retrieved from https://www.kiro7.com/news/local/anti-trump-protesters-gather-again-in-seattle/489087028

Austin, J. L. (1975). *How to do things with words*: Oxford university press.

Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against muslims on social media. *Policy & Internet, 6*(2), 133-150.

Bacon, J., & Gomez, A. (2017). Protests against Trump's immigration plan rolling in more than 30 cities. *USA Today*. Retrieved from https://www.usatoday.com/story/news/nation/2017/01/29/homeland-security-judges-stay-has-little-impact-travel-ban/97211720/

Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). *Deep learning for hate speech detection in tweets.* Paper presented at the Proceedings of the 26th International Conference on World Wide Web Companion.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132.

Balgopal, P. R. (2000). *Social work practice with immigrants and refugees*. New York: Columbia University Press.

Banaszak, L. A., & Ondercin, H. L. (2016). Public Opinion as a Movement Outcome: The Case of the US Women's Movement. *Mobilization: An International Quarterly, 21*(3), 361-378.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., . . . Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin, 38*(12), 1629-1643.

Ben-David, A., & Matamoros-Fernandez, A. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication, 10*, 1167-1193.

Bernstein, M. (2003). Nothing ventured, nothing gained? Conceptualizing social movement "success" in the lesbian and gay movement. *Sociological Perspectives, 46*(3), 353-379.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory explaining and reducing children's social stereotyping and prejudice. *Current directions in psychological science, 16*(3), 162-166.

Birtel, M. D., & Crisp, R. J. (2015). Psychotherapy and Social Change: Utilizing Principles of Cognitive-Behavioral Therapy to Help Develop New Prejudice-Reduction Interventions. *Frontiers in psychology, 6*, 1771.

Blackledge, A. (2005). *Discourse and power in a multilingual world* (Vol. 15): John Benjamins Publishing.

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology, 79*(6), 993.

Blaya, C. (2019). Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior, 45*, 163-172.

Bobo, L. (1997). Race, public opinion, and the social sphere. *The Public Opinion Quarterly, 61*(1), 1-15.

Bolton, E. B. (1935). Effect of knowledge upon attitudes towards the negro. *The Journal of Social Psychology, 6*(1), 68-90.

Bonilla, Y., & Rosa, J. (2015). # Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist, 42*(1), 4-17.

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology, 38*(3), 551-569.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*: Oxford University Press.

Boyle, M. P., Dioguardi, L., & Pate, J. E. (2016). A comparison of three strategies for reducing the public stigma associated with stuttering. *Journal of fluency disorders, 50*, 44-58.

Brady, S. R., Young, J. A., & McLeod, D. A. (2015). Utilizing digital advocacy in community organizing: Lessons learned from organizing in virtual spaces to promote worker rights and economic justice. *Journal of Community Practice, 23*(2), 255-273.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313-7318.

Branton, R., Martinez-Ebers, V., Carey, T. E., & Matsubayashi, T. (2015). Social protest and policy attitudes: The case of the 2006 immigrant rallies. *American Journal of Political Science, 59*(2), 390-402.

Breitzer, S. R. (2011). Race, Immigration, and Contested Americanness: Black Nativism and the American Labor Movement, 1880-1930. *Race/Ethnicity: Multidisciplinary Global Contexts, 4*(2), 269-283.

Brewer, M. B., & Miller, N. (1984). *Groups in contact: The psychology of desegregation*: Academic Press.

Britt, T. W., Bonieci, K. A., Vescio, T. K., Biernat, M., & Brown, L. M. (1996). Intergroup anxiety: A person× situation approach. *Personality and Social Psychology Bulletin, 22*(11), 1177-1188.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science, 352*(6282), 220-224.

Brown, A., & Stepler, R. (2016). Statistical Portrait of the Foreign-Born Population in the United States. Retrieved from http://www.pewhispanic.org/2016/04/19/statistical-portrait-of-the-foreign-born-population-in-the-united-states-key-charts/ - 2013-fb-first-second-gen

Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social Forces, 55*(4), 881-897.

Burnap, P., & Williams, M. L. (2014). Hate speech, machine classification and statistical modelling of information flows on Twitter: Interpretation and communication for policy decision making.

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet, 7*(2), 223-242.

Burnap, P., & Williams, M. L. (2016). Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science, 5*(1), 1.

Burstein, P. (1999). Sociol Movements ond Public Policy. *How social movements matter, 10*, 1.

Burstein, P., & Freudenburg, W. (1978). Changing public policy: The impact of public opinion, antiwar demonstrations, and war costs on senate voting on Vietnam war motions. *American journal of sociology, 84*(1), 99-122.

Cakal, H., Hewstone, M., Schwär, G., & Heath, A. (2011). An investigation of the social identity model of collective action and the 'sedative'effect of intergroup contact among Black and White students in South Africa. *British journal of social psychology, 50*(4), 606-627.

California Senate Bill No. 670. (2005). Apology Act for the 1930s Mexican Repatriation Program.

Carey Jr, T. E., Branton, R. P., & Martinez-Ebers, V. (2014). The Influence of Social Protests on Issue Salience among Latinos. *Political Research Quarterly, 67*(3), 615-627.

Caricati, L., Mancini, T., & Marletta, G. (2017). The role of ingroup threat and conservative ideologies on prejudice against immigrants in two samples of Italian adults. *The Journal of Social Psychology, 157*(1), 86-97.

Casados, A. T. (2017). Reducing the stigma of mental illness: current approaches and future directions. *Clinical Psychology: Science and Practice, 24*(3), 306-323.

Cervantes, R. C., Padilla, A. M., Napper, L. E., & Goldbach, J. T. (2013). Acculturation-related stress and mental health outcomes among three generations of Hispanic adolescents. *Hispanic Journal of Behavioral Sciences*, 0739986313500924.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science, 24*(2), 140-149.

Chan, J., Ghose, A., & Seamans, R. (2016). The Internet and Racial Hate Crime: Offline Spillovers from Online Access. *MIS Quarterly, 40*(2), 381-403.

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Mean Birds: Detecting Aggression and Bullying on Twitter. *arXiv preprint arXiv:1702.06877*.

Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies, 65*(1), 57-70.

Chenoweth, E., Stephan, M. J., & Stephan, M. J. (2011). *Why civil resistance works: The strategic logic of nonviolent conflict*: Columbia University Press.

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday, 11*(11).

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one, 5*(11), e14118.

Chiluwa, I., & Ifukor, P. (2015). 'War against our Children': Stance and evaluation in# BringBackOurGirls campaign discourse on Twitter and Facebook. *Discourse & society, 26*(3), 267-296.

Chung, W., He, S., Zeng, D. D., & Benjamin, V. (2015). *Emotion extraction and entrainment in social media: The case of US immigration and border security.* Paper presented at the Intelligence and Security Informatics (ISI), 2015 IEEE International Conference.

Chung, W., Wei, K., Yu, Y., & Wen, X. (2016). *The Dynamics of Group Risk Perception in the US After Paris Attacks.* Paper presented at the Social Informatics: 8th International Conference, SocInfo 2016, Seattle, US.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in experimental social psychology, 24*, 201-234.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd: Hillsdale, NJ: erlbaum.

Cohrs, J. C., & Stelzl, M. (2010). How ideological attitudes predict host society members' attitudes toward immigrants: Exploring cross - national differences. *Journal of social issues, 66*(4), 673-694.

Comas-Forgas, R., Sureda-Negre, J., & Calvo-Sastre, A. (2017). Characteristics of Cyberbullying Among Native and Immigrant Secondary Education Students. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), 7*(1), 1-17.

Corrigan, P. W., Morris, S. B., Michaels, P. J., Rafacz, J. D., & Rüsch, N. (2012). Challenging the public stigma of mental illness: a meta-analysis of outcome studies. *Psychiatric services, 63*(10), 963-973.

Corrigan, P. W., & Penn, D. L. (1999). Lessons from social psychology on discrediting psychiatric stigma. *American Psychologist, 54*(9), 765.

Corrigan, P. W., Roe, D., & Tsang, H. W. (2011). *Challenging the stigma of mental illness: Lessons for therapists and advocates*: John Wiley & Sons.

Corrigan, P. W., & Watson, A. C. (2002). The paradox of self - stigma and mental illness. *Clinical Psychology: Science and Practice, 9*(1), 35-53.

Coser, L. A. (1957). Social conflict and the theory of social change. *The British Journal of Sociology, 8*(3), 197-207.

Cottrell, C., & Neuberg, S. (2003). *From patterns of threat to patterns of behavior: Capturing the complexity of intergroup interaction.* Paper presented at the annual meeting of the Society for Personality and Social Psychology, Los Angeles.

Cottrell, C. A., & Neuberg, S. L. (2005). Different Emotional Reactions to Different Groups: A Sociofunctional Threat-Based Approach to "Prejudice". *Journal of Personality and Social Psychology, 88*(5), 770-789. doi:10.1037/0022-3514.88.5.770

Crandall, C. S., Eshleman, A., & O'brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359.

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009*.

De Choudhury, M., Jhaver, S., Sugar, B., & Weber, I. (2016). *Social Media Participation in an Activist Movement for Racial Equality.* Paper presented at the ICWSM.

De Cillia, R., Reisigl, M., & Wodak, R. (1999). The discursive construction of national identities. *Discourse & society, 10*(2), 149-173.

DeFriese, G. H., & Ford, W. S. (1969). Verbal attitudes, overt acts, and the influence of social constraint in interracial behavior. *Social Problems, 16*(4), 493-505.

Demick, B. (2017). Thousands at JFK airport in New York protest new visa and refugee suspensions. *Los Angeles Times*. Retrieved from http://www.latimes.com/politics/washington/la-na-trailguide-updates-thousands-at-jfk-airport-in-new-york-1485642642-htmlstory.html

DHS. (2003). ENDGAME Office of Detention and Removal Strategic Plan, 2003 - 2012 Detention and Removal Strategy for a Secure Homeland. Retrieved from http://fas.org/irp/agency/dhs/endgame.pdf

Dijker, A. J. (1987). Emotional reactions to ethnic minorities. *European Journal of Social Psychology, 17*(3), 305-325.

Dixon, J., Durrheim, K., Kerr, P., & Thomae, M. (2013). 'What's So Funny'Bout Peace, Love and Understanding?'Further Reflections on the Limits of Prejudice Reduction as a Model of Social Change. *Journal of Social and Political Psychology, 1*(1), 239-252.

Dixon, J., Durrheim, K., Stevenson, C., & Cakal, H. (2016). From prejudice reduction to collective action: Two psychological models of social change (and how to reconcile them). *Cambridge Handbook of the Psychology of Prejudice. Cambridge: Cambridge University Press. contact prevalence in five Central European countries. European Journal of Social Psychology, 44*(6), 536-547.

Dixon, J., Levine, M., Reicher, S., & Durrheim, K. (2012). Beyond prejudice: Are negative evaluations the problem and is getting us to like one another more the solution? *Behavioral and Brain Sciences, 35*(06), 411-425.

Dixon, J., Tropp, L. R., Durrheim, K., & Tredoux, C. (2010). “Let them eat harmony” prejudice-reduction strategies and attitudes of historically disadvantaged groups. *Current directions in psychological science, 19*(2), 76-80.

Dred Scott v. Sandford, 60 US 393 C.F.R. (1857).

Dru, V. (2007). Authoritarianism, social dominance orientation and prejudice: Effects of various self-categorization conditions. *Journal of Experimental Social Psychology, 43*(6), 877-883.

Duckitt, J., & Sibley, C. G. (2007). Right wing authoritarianism, social dominance orientation and the dimensions of generalized prejudice. *European Journal of Personality, 21*(2), 113-130.

Dumont, M., Yzerbyt, V., Wigboldus, D., & Gordijn, E. H. (2003). Social categorization and fear reactions to the September 11th terrorist attacks. *Personality and Social Psychology Bulletin, 29*(12), 1509-1520.

Echebarria ‐ Echabe, A., & Guede, E. F. (2007). A new measure of anti ‐ Arab prejudice: Reliability and validity evidence. *Journal of Applied Social Psychology, 37*(5), 1077-1091.

Edwards, C. (2017). 3K protesters descend on LAX for 2nd day in opposition of Trump travel ban. *ABC*. Retrieved from https://abc7.com/news/3k-protesters-descend-on-lax-in-opposition-to-trumps-travel-ban/1726960/

Ekehammar, B., & Akrami, N. (2003). The relation between personality and prejudice: a variable ‐ and a person ‐ centred approach. *European Journal of Personality, 17*(6), 449-464.

Ellemers, N., & Barreto, M. (2009). Collective action in modern times: How modern expressions of prejudice prevent collective action. *Journal of social issues, 65*(4), 749-768.

ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018). Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. *arXiv preprint arXiv:1804.04257*.

Epps, D. J. (2015). The HEAR.US Project-Reducing anti-immigrant sentiment and myth through an online awareness intervention. Retrieved from http://digitalcommons.tacoma.uw.edu/msw_capstones/18/

Esses, V. M., Wagner, U., Wolf, C., Preiser, M., & Wilbur, C. J. (2006). Perceptions of national identity and attitudes toward immigrants and immigration in Canada and Germany. *International Journal of Intercultural Relations, 30*(6), 653-669.

Falomir-Pichastor, J. M., & Frederic, N. S. (2013). The dark side of heterogeneous ingroup identities: National identification, perceived threat, and prejudice against immigrants. *Journal of Experimental Social Psychology, 49*(1), 72-79.

Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger is more influential than joy: Sentiment correlation in Weibo. *PloS one, 9*(10), e110184.

Fangen, K. (2010). Social exclusion and inclusion of young immigrants: Presentation of an analytical framework. *Young, 18*(2), 133-156.

FBI. (2001). Uniform Crime Reports. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2001

Feddes, A., Noack, P., & Rutland, A. (2009). Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study. *Child development, 80*(2), 377-390.

Ferguson, E. E. (1947). The California Alien Land Law and the Fourteenth Amendment. *Cal. L. Rev., 35*, 61.

Firzsimon, C. (2007). Hispanic people feel new hositility. Retrieved from http://www.ncpolicywatch.com/2007/09/23/hispanic-people-feel-new-hostility/

Fisher, E. L., Deason, G., Borgida, E., & Oyamot, C. M. (2011). A model of authoritarianism, social norms, and personal values: Implications for Arizona law enforcement and immigration policy. *Analyses of Social Issues and Public Policy, 11*(1), 285-299.

Flores-Yeffal, N. Y., Vidales, G., & Plemons, A. (2011). The Latino cyber-moral panic process in the United States. *Information, Communication & Society, 14*(4), 568-589.

Florido, A. (2015). Tracing The Shifting Meaning Of 'Alien'. Retrieved from http://www.npr.org/sections/codeswitch/2015/08/22/432774244/tracing-the-shifting-meaning-of-alien

França, D. X., & Monteiro, M. B. (2013). Social norms and the expression of prejudice: The development of aversive racism in childhood. *European Journal of Social Psychology, 43*(4), 263-271.

Fromm, E. (1941). *Escape from freedom*. New York City, NY: Farrar & Rinehart.

Furman, R., Langer, C. L., Sanchez, T. W., & Negi, N. J. (2007). A qualitative study of immigration policy and practice dilemmas for social work students. *Journal of Social Work Education, 43*(1), 133-146.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. UK: Psychology Press.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57*(2), 239.

Gamson, W. (1975). *The Strategy of Social Protest*. Belmont, Cal.: Wadsworth.

Gandy-Guedes, M. E., Vance, M. M., Bridgewater, E. A., Montgomery, T., & Taylor, K. (2016). Using Facebook as a tool for informal peer support: a case example. *Social Work Education, 35*(3), 323-332.

Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer‐Mediated Communication, 17*(3), 319-336.

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering, 10*(4), 215-230.

Giugni, M. (2004). *Social protest and policy change: Ecology, antinuclear, and peace movements in comparative perspective*: Rowman & Littlefield.

Golbeck, J., Ashktorab, Z., Banjo, R. O., Berlinger, A., Bhagwan, S., Buntain, C., . . . Gnanasekaran, R. K. (2017). *A Large Labeled Corpus for Online Harassment Research.* Paper presented at the Proceedings of the 2017 ACM on Web Science Conference.

Goldenberg, A., Halperin, E., van Zomeren, M., & Gross, J. J. (2016). The Process Model of Group-Based Emotion: Integrating Intergroup Emotion and Emotion Regulation Perspectives. *Personality and Social Psychology Review, 20*(2), 118-141. doi:10.1177/1088868315581263

Goldman, L. M. (2015). Trending now: the use of social media websites in public shaming punishments. *Am. Crim. L. Rev., 52*, 415.

Google. (2019). Inappropriate content. Retrieved from https://support.google.com/merchants/answer/6149990?hl=en

Gordijn, E. H., Yzerbyt, V., Wigboldus, D., & Dumont, M. (2006). Emotional reactions to harmful intergroup behavior. *European Journal of Social Psychology, 36*(1), 15-30.

Graham, D. (2016). Trump: Mexican American Judge Has an 'Absolute Conflict'. Retrieved from https://www.theatlantic.com/politics/archive/2016/06/trump-mexican-judge/485429/

Graumann, C. F., & Wintermantel, M. (1989). Discriminatory speech acts: A functional approach. *Stereotyping and prejudice: Changing conceptions*, 184-204.

Greenwood, S., Perrin, A., & Duggan, M. (2016). Social Media Update 2016 Facebook usage and engagement is on the rise, while adoption of other platforms holds steady. Retrieved from http://www.pewinternet.org/2016/11/11/social-media-update-2016/

Greeson, J. K., An, S., Xue, J., Thompson, A. E., & Guo, C. (2018). Tweeting social justice: How social work faculty use Twitter. *The British Journal of Social Work, 48*(7), 2038-2057.

Grimmer, J. (2015). We are all social scientists now: how big data, machine learning, and causal inference work together. *PS: Political Science & Politics, 48*(1), 80-83.

Guberman, J., Schmitz, C., & Hemphill, L. (2016). *Quantifying toxicity and verbal violence on Twitter.* Paper presented at the Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences, 18*, 43-49.

The Border Protection, Anti-terrorism, and Illegal Immigration Control Act of 2005, H.R. 4437 109th Congress C.F.R. (2005).

Habryn, F. (2014). *Customer intimacy analytics: Leveraging operational data to assess customer knowledge and relationships and to measure their business impact*: KIT Scientific Publishing.

Hamblin, R. L. (1962). The dynamics of racial discrimination. *Social Problems, 10*(2), 103-121.

Hanna, A.-M. V., & Ortega, D. M. (2016). Salir adelante (perseverance): Lessons from the Mexican immigrant experience. *Journal of Social Work, 16*(1), 47-65. doi:10.1177/1468017314560301

Harlow, S., & Guo, L. (2014). Will the revolution be tweeted or facebooked? Using digital communication tools in immigrant activism. *Journal of Computer‐Mediated Communication, 19*(3), 463-478.

Hartley, R. D., & Armendariz, L. F. (2011). Border justice? Sentencing federal narcotics offenders in southwest border districts: A focus on citizenship status. *Journal of contemporary criminal justice, 27*(1), 43-62.

Haslam, N. (2012). What's so crummy'bout peace, love, and understanding? *Behavioral and Brain Sciences, 35*(06), 434-435.

He, X., Lu, D., Margolin, D., Wang, M., Idrissi, S. E., & Lin, Y.-R. (2017). *The Signals and Noise: Actionable Information in Improvised Social Media Channels During a Disaster.* Paper presented at the Proceedings of the 2017 ACM on Web Science Conference.

Hiel, A. V., & Mervielde, I. (2005). Authoritarianism and social dominance orientation: Relationships with various forms of racism. *Journal of Applied Social Psychology, 35*(11), 2323-2344.

Higham, J. (2002). *Strangers in the land: Patterns of American nativism, 1860-1925*. New Brunswick, NJ: Rutgers University Press.

Hitchcock, L. I., & Young, J. A. (2016). Tweet, tweet!: Using live Twitter chats in social work education. *Social Work Education, 35*(4), 457-468.

Hite, L. M., & Mc Donald, K. S. (2006). Diversity training pitfalls and possibilities: An exploration of small and mid-size US organizations. *Human Resource Development International, 9*(3), 365-377.

Hitlan, R. T., Carrillo, K., Zárate, M. A., & Aikman, S. N. (2007). Attitudes toward immigrant groups and the September 11 terrorist attacks. *Peace and Conflict: Journal of Peace Psychology, 13*(2), 135.

Hodson, G., & Esses, V. M. (2005). Lay perceptions of ethnic prejudice: Causes, solutions, and individual differences. *European Journal of Social Psychology, 35*(3), 329-344.

Horn, B. (2017). Thousands at Philly airport call Trump's order 'un-American'. *Delaware Online*. Retrieved from https://www.delawareonline.com/story/news/local/2017/01/29/delaware-immigration-trump-protest/97215226/

Imhoff, R., & Recker, J. (2012). Differentiating Islamophobia: Introducing a new scale to measure Islamoprejudice and secular Islam critique. *Political Psychology, 33*(6), 811-824.

Islam, M. R., & Hewstone, M. (1993). Dimensions of contact as predictors of intergroup anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality and Social Psychology Bulletin, 19*(6), 700-710.

Jiménez, T. R. (2007). Weighing the Costs and Benefits of Mexican Immigration: The Mexican - American Perspective. *Social Science Quarterly, 88*(3), 599-618.

Jones, R. H. (2015). Generic intertextuality in online social activism: The case of the It Gets Better project. *Language in Society, 44*(3), 317-339.

Jost, J. T., Stern, C., Rule, N. O., & Sterling, J. (2017). The Politics of Fear: Is There an Ideological Asymmetry in Existential Motivation? *Social Cognition, 35*(4), 324.

Kaiser Family Foundation. (2013). Key Facts on Health Coverage for Low-Income Immigrants Today and Under the Affordable Care Act. Retrieved from https://kaiserfamilyfoundation.files.wordpress.com/2013/03/8279-02.pdf

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American sociological review, 71*(4), 589-617.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology, 28*(3), 280.

Kinder, D. R., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*: University of Chicago Press.

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology, 40*(3), 414.

King, D. S., & Smith, R. M. (2005). Racial orders in American political development. *American Political Science Review, 99*(1), 75-92.

Kirst-Ashman, K. K., & Hull, G. H., Jr. (2002). *Understanding generalist practice* (3rd ed.). Pacific Grove, CA: Brooks/Cole Thomson Learning.

Kitzinger, J. (1995). Qualitative research: introducing focus groups. *Bmj, 311*(7000), 299-302.

Klöckner, C. A., & de Raaf, S. (2013). Expression of prejudice against immigrants in a group situation: the impact of context, attitudes, and egalitarian values. *SAGE Open, 3*(3), 2158244013504574.

Kluegel, J. R., & Smith, E. R. (1986). *Beliefs about inequality: Americans' views of what is and what ought to be*: Transaction Publishers.

Kohn, S. (2015). Trump's outrageous Mexico remarks. Retrieved from http://www.cnn.com/2015/06/17/opinions/kohn-donald-trump-announcement/.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological bulletin, 140*(4), 1073-1137.

Krogstad, J. (2015). Social media preferences vary by race and ethnicity. Retrieved from http://www.pewresearch.org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity/

Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another Look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative Differences in" Moral" Intuitions? *Social Justice Research, 27*(4), 413.

Kuppens, T., & Yzerbyt, V. Y. (2012). Group-based emotions: The impact of social identity on appraisals, emotions, and behaviors. *Basic and Applied Social Psychology, 34*(1), 20-33.

Kuppens, T., Yzerbyt, V. Y., Dandache, S., Fischer, A. H., & van der Schalk, J. (2013). Social identity salience shapes group-based emotions through group-based appraisals. *Cognition & Emotion, 27*(8), 1359-1377.

Kwok, I., & Wang, Y. (2013). *Locate the Hate: Detecting Tweets against Blacks.* Paper presented at the AAAI.

Lakoff, G. (1997). *Moral politics: What conservatives know that liberals don't*. Chicago: University Of Chicago Press.

Lane, R. E. (1962). Political Ideology: Why the Common Man Believes What He Does about Politics: Glencoe, IL: Free Press.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in human behavior, 28*(2), 434-443.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist, 37*(9), 1019.

Lazarus, R. S. (1991). *Emotion and adaptation*: Oxford University Press on Demand.

Leader, T., Mullen, B., & Rice, D. (2009). Complexity and valence in ethnophaulisms and exclusion of ethnic out-groups: what puts the" hate" into hate speech? *Journal of Personality and Social Psychology, 96*(1), 170.

Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist, 45*(6), 927-957.

Leonard, D. J., Mackie, D. M., & Smith, E. R. (2011). Emotional responses to intergroup apology mediate intergroup forgiveness and retribution. *Journal of Experimental Social Psychology, 47*(6), 1198-1206.

Leonard, D. J., Moons, W. G., Mackie, D. M., & Smith, E. R. (2011). "We're mad as hell and we're not going to take it anymore": Anger self-stereotyping and collective action. *Group Processes & Intergroup Relations, 14*(1), 99-111.

Levy, B. R., Chung, P. H., Bedford, T., & Navrazhina, K. (2013). Facebook as a site for negative age stereotypes. *The Gerontologist, 54*(2), 172-176.

Lewin, K. (1947). Group decision and social change. *Readings in social psychology, 3*, 197-211.

Lichtablau, E. (2015). Crimes Against Muslim Americans and Mosques Rise Sharply. *The New York Times*. Retrieved from http://www.nytimes.com/2015/12/18/us/politics/crimes-against-muslim-americans-and-mosques-rise-sharply.html?smid=tw-nytimes&smtyp=cur&_r=0

Light, M. T., Massoglia, M., & King, R. D. (2014). Citizenship and punishment: The salience of national membership in US Criminal Courts. *American sociological review, 79*(5), 825-847.

Lin, Y.-R., Keegan, B., Margolin, D., & Lazer, D. (2014). Rising tides or rising stars?: Dynamics of shared attention on Twitter during media events. *PloS one, 9*(5), e94093.

Lin, Y.-R., Margolin, D., Keegan, B., & Lazer, D. (2013). *Voices of victory: A computational focus group framework for tracking opinion shift in real time.* Paper presented at the Proceedings of the 22nd international conference on World Wide Web.

Lin, Y. R., Margolin, D., & Wen, X. (2017). Tracking and analyzing individual distress following terrorist attacks using social media streams. *Risk analysis, 37*(8), 1580-1605.

Lippitt, R. (1949). *Training in community relations*. New York: Harper.

López, M. P. (2012). Essay Examining the Murder of Luis Ramirez and the Emergence of Hate Crimes against Latino Immigrants in the United States, An. *Ariz. St. LJ, 44*, 155.

López Peláez, A., Pérez García, R., & Aguilar-Tablada Massó, M. V. (2018). e-Social work: building a new field of specialization in social work? *European Journal of Social Work, 21*(6), 804-823.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*(5), 842.

Lozano, E., Cedeño, J., Castillo, G., Layedra, F., Lasso, H., & Vaca, C. (2017). *Requiem for online harassers: Identifying racism from political tweets.* Paper presented at the eDemocracy & eGovernment (ICEDEG), 2017 Fourth International Conference on.

Luty, J., Umoh, O., Sessay, M., & Sarkhel, A. (2007). Effectiveness of Changing Minds campaign factsheets in reducing stigmatised attitudes towards mental illness. *Psychiatric Bulletin, 31*(10), 377-381.

Mackie, D., & Smith, E. (2015). Intergroup Emotions. In M. Mikulincer & P. Shaver (Eds.), *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences* (Vol. 4). Washington, D.C.: American Psychological Association.

Mackie, D. M., Maitner, A. T., & Smith, E. R. (2009). Intergroup emotions theory. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 285-307). Abingdon, United Kingdom: Psychology Press.

MacWilliams, M. C. (2016). Who decides when the party doesn't? Authoritarian voters and the rise of Donald Trump. *PS: Political Science & Politics, 49*(4), 716-721.

Madestam, A., Shoag, D., Veuger, S., & Yanagizawa-Drott, D. (2013). Do political protests matter? Evidence from the tea party movement. *The Quarterly Journal of Economics, 128*(4), 1633-1685.

Magdy, W., Darwish, K., & Abokhodair, N. (2015). Quantifying Public Response towards Islam on Twitter after Paris Attacks. *arXiv preprint arXiv:1512.04570*.

Maitner, A. T., Mackie, D. M., & Smith, E. R. (2006). Evidence for the regulatory function of intergroup emotion: Emotional consequences of implemented or impeded intergroup action tendencies. *Journal of Experimental Social Psychology, 42*(6), 720-728.

Manjoo, F. (2017). The Alt-Majority: How Social Networks Empowered Mass Protests Against Trump. *New York Times*. Retrieved from https://www.nytimes.com/2017/01/30/technology/donald-trump-social-networks-protests.html

Marrow, A. J., & French, J. R. (1945). Changing a stereotype in industry. *Journal of social issues, 1*(3), 33-37.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review, 50*(4), 370.

Mastnak, T. (2008). *Western hostility toward the Muslims: a history of the present.* Paper presented at the Forum Bosnae.

Mazumder, S. (2018). The Persistent Effect of US Civil Rights Protests on Political Attitudes. *American Journal of Political Science, 62*(4), 922-935.

McCammon, H. J., Chaudhuri, S., Hewitt, L., Muse, C. S., Newman, H. D., Smith, C. L., & Terrell, T. M. (2008). Becoming full citizens: The US women's jury rights campaigns, the pace of reform, and strategic adaptation. *American journal of sociology, 113*(4), 1104-1147.

McClosky, H. (1958). Conservatism and Personality. *American Political Science Review, 52*(01), 27-45.

McCorriston, J., Jurgens, D., & Ruths, D. (2015). *Organizations Are Users Too: Characterizing and Detecting the Presence of Organizations on Twitter.* Paper presented at the ICWSM.

McNamee, L. G., Peterson, B. L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs, 77*(2), 257-280.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, D. A., Smith, E. R., & Mackie, D. M. (2004). Effects of intergroup contact and political predispositions on prejudice: Role of intergroup emotions. *Group Processes & Intergroup Relations, 7*(3), 221-237.

Miller, L. W., Polinard, J. L., & Wrinkle, R. D. (1984). Attitudes toward undocumented workers: the Mexican American perspective. *Social Science Quarterly, 65*(2), 482-494.

Moon, A., & Fares, M. (2016). Trump's call to ban Muslims draws fire on social media. Retrieved from http://www.reuters.com/article/us-usa-election-trump-socialmedia-idUSKBN0TR2K020151208

Moshe Grodofsky, M., & Makaros, A. (2016). Social Work and Social Protest: An Under-Researched Field. *Journal of Policy Practice, 15*(4), 333-348.

Mukherjee, S., Molina, L. E., & Adams, G. (2012). National identity and immigration policy: Concern for legality or ethnocentric exclusion? *Analyses of Social Issues and Public Policy, 12*(1), 21-32.

Mullen, B., & Rice, D. R. (2003). Ethnophaulisms and exclusion: The behavioral consequences of cognitive representation of ethnic immigrant groups. *Personality and Social Psychology Bulletin, 29*(8), 1056-1067.

Murray, K. E., & Marx, D. M. (2013). Attitudes toward unauthorized immigrants, authorized immigrants, and refugees. *Cultural Diversity and Ethnic Minority Psychology, 19*(3), 332.

NAMI. (2000). NAMI protests ABC’s Wonderland commercial sponsors. Retrieved from https://www.nami.org/Press-Media/Press-Releases/2000/NAMI-Protests-ABC%E2%80%99s-Wonderland-With-Commercial-Spo

NASW. (2017). NASW STATEMENT: President Trump’s immigration executive order is inhumane, de facto ban on Muslim immigrants. Retrieved from http://www.socialworkblog.org/advocacy/2017/01/nasw-statement-president-trumps-immigration-executive-order-is-inhumane-de-facto-ban-on-muslim-immigrants/

Ng, W. L. (2002). *Japanese American internment during World War II: A history and reference guide*: Greenwood Publishing Group.

Nier, J. A., Gaertner, S. L., Nier, C. L., & Dovidio, J. F. (2012). Can racial profiling be avoided under Arizona immigration law? Lessons learned from subtle bias research and anti‐discrimination law. *Analyses of Social Issues and Public Policy, 12*(1), 5-20.

Oentaryo, R. J., Low, J.-W., & Lim, E.-P. (2015). *Chalk and cheese in twitter: Discriminating personal and organization accounts.* Paper presented at the European Conference on Information Retrieval.

Olteanu, A., Castillo, C., Boy, J., & Varshney, K. R. (2018). The Effect of Extremist Violence on Hateful Speech Online. *arXiv preprint arXiv:1804.05704*.

Oswald, D. L. (2005). Understanding anti‐arab reactions post‐9/11: the role of threats, social categories, and personal ideologies. *Journal of Applied Social Psychology, 35*(9), 1775-1799.

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *Cyberpsychology & behavior, 12*(6), 729-733.

Perkins, D. M., & Mebert, C. J. (2005). Efficacy of Multicultural Education for Preschool Children A Domain-Specific Approach. *Journal of cross-cultural psychology, 36*(4), 497-512.

Pettigrew, T. F. (1958). Personality and sociocultural factors in intergroup attitudes: A cross-national comparison. *Journal of conflict resolution, 2*(1), 29-42.

Pettigrew, T. F. (1959). Regional differences in anti-Negro prejudice. *The Journal of Abnormal and Social Psychology, 59*(1), 28.

Pettigrew, T. F. (1960). Social distance attitudes of South African students. *Social Forces, 38*(3), 246-253.

Pettigrew, T. F. (2017). Social Psychological Perspectives on Trump Supporters. *Journal of Social and Political Psychology, 5*(1), 107-116.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751.

Pew Research Center. (2017). Key findings about US immigrants. Retrieved from http://www.pewresearch.org/fact-tank/2017/05/03/key-findings-about-u-s-immigrants/

Pew Research Center. (2018). Social Media Fact Sheet. Retrieved from http://www.pewinternet.org/fact-sheet/social-media/

Phelan, J. C., Link, B. G., & Dovidio, J. F. (2008). Stigma and prejudice: one animal or two? *Social science & medicine, 67*(3), 358-367.

Piven, F. F., & Cloward, R. (1977). *Poor people's movements: Why they succeed, how they fail*: Vintage.

Pottie‐Sherman, Y., & Wilkes, R. (2017). Does size really matter? On the relationship between immigrant group size and anti‐immigrant prejudice. *International migration review, 51*(1), 218-250.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741.

Chinese Exclusion Act, 22 Stat. 58 C.F.R. (1882).

Public Law 111-148. (2010). Protection Patient Act and Affordable Care,. *Title IV, x4207, USC HR, 3590*, 2010.

Quimby, B. (2017). Thousands protest Trump policies at 2 rallies in Portland. *Press herald*. Retrieved from https://www.pressherald.com/2017/01/29/anti-trump-rallies-are-popping-up-around-maine-today/

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta ‐ analysis of age differences. *Child development, 82*(6), 1715-1737.

Rainie, L., Smith, A., Schlozman, K. L., Brady, H., & Verba, S. (2012). Social media and political engagement. *Pew Internet & American Life Project, 19*, 2-13.

Rana, J. (2007). The story of Islamophobia. *Souls, 9*(2), 148-161.

Ray, D. G., Mackie, D., & Smith, E. R. (2014). Intergroup emotion: self-categorization, emotion, and the regulation of intergroup conflict. In C. von Scheve & M. Salmeda (Eds.), *Collective Emotions*. Oxford, UK:: Oxford University Press.

Ray, D. G., Mackie, D. M., Rydell, R. J., & Smith, E. R. (2008). Changing categorization of self can change emotions about outgroups. *Journal of Experimental Social Psychology, 44*(4), 1210-1213.

Ray, D. G., Mackie, D. M., Smith, E. R., & Terman, A. W. (2012). Discrete emotions elucidate the effects of crossed-categorization on prejudice. *Journal of Experimental Social Psychology, 48*(1), 55-69.

Roblain, A., Malki, B., Azzi, A., & Licata, L. (2017). After coming in, settling in: An analysis of early-stage acculturation preferences of male Syrian and Iraqi asylum seekers in Belgium. *International Review of Social Psychology, 30*(1).

Rojas, H., & Puig‐i‐Abril, E. (2009). Mobilizers mobilized: Information, expression, mobilization and participation in the digital age. *Journal of Computer‐Mediated Communication, 14*(4), 902-927.

Romero, F. (2008). *Hyperborder: the contemporary US-Mexico border and its future*. New York City, NY: Princeton Architectural Press.

Romero, M. (2006). Racial profiling and immigration law enforcement: Rounding up of usual suspects in the Latino community. *Critical Sociology, 32*(2-3), 447-473.

Rourke, M. (2017). Twitter lost 2 million users in the US last quarter. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2017/07/27/twitter-lost-2-million-users-in-the-u-s-last-quarter/?utm_term=.a65981043f42

Rubin, J. D., & McClelland, S. I. (2015). 'Even though it's a small checkbox, it's a big deal': stresses and strains of managing sexual identity (s) on Facebook. *Culture, health & sexuality, 17*(4), 512-526.

Rüsch, N., Angermeyer, M. C., & Corrigan, P. W. (2005). Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. *European Psychiatry, 20*(8), 529-539.

Rydell, R. J., Mackie, D. M., Maitner, A. T., Claypool, H. M., Ryan, M. J., & Smith, E. R. (2008). Arousal, processing, and risk taking: Consequences of intergroup anger. *Personality and Social Psychology Bulletin, 34*(8), 1141-1152.

Saguy, T., Tausch, N., Dovidio, J. F., & Pratto, F. (2009). The irony of harmony: Intergroup contact can produce false expectations for equality. *Psychological Science, 20*(1), 114-121.

Santa Ana, O. (1999). Like an animal I was treated': Anti-immigrant metaphor in US public discourse. *Discourse & society, 10*(2), 191-224.

Sayre, B., Bode, L., Shah, D., Wilcox, D., & Shah, C. (2010). Agenda setting in a digital age: Tracking attention to California Proposition 8 in social media, online news and conventional news. *Policy & Internet, 2*(2), 7-32.

Schäffner, C. (2002). *The role of discourse analysis for translation and in translator training*: Multilingual matters.

Schuman, H. (1997). *Racial attitudes in America: trends and interpretations* (Rev. ed.). Cambridge, Mass: Harvard University Press.

Schwartz, A. (1999). Americanization and cultural preservation in Seattle's Settlement House: A Jewish adaptation of the Anglo-American model of settlement work. *J. Soc. & Soc. Welfare, 26*, 25.

Sears, D. O. (1988). Eliminating racism. New York City, NY: Springer.

Seger, C. R., Smith, E. R., & Mackie, D. M. (2009). Subtle activation of a social categorization triggers group-level emotions. *Journal of Experimental Social Psychology, 45*(3), 460-467.

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review, 14*(3), 197-215.

Shaffer, L. S. (1983). Toward Pepitone's vision of a normative social psychology: What is a social norm? *The journal of mind and behavior*, 275-293.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of misinformation by social bots: e-print.

Sheehi, S. (2011). *Islamophobia: The ideological campaign against Muslims*. Atlanta, GA: Clarity Press.

Sherif, M., & Sherif, C. W. (1953). Groups in harmony and tension; an integration of studies of intergroup relations.

Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign affairs*, 28-41.

Short, R., & Magaña, L. (2002). Political Rhetoric, Immigration Attitudes, and Contemporary Prejudice: A Mexican American Dilemma. *The Journal of Social Psychology, 142*(6), 701-712. doi:10.1080/00224540209603930

Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*: Cambridge University Press.

Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology, 70*(3), 476.

Sidanius, J., Pratto, F., Martin, M., & Stallworth, L. M. (1991). Consensual racism and career track: Some implications of social dominance theory. *Political Psychology*, 691-721.

Siegel, S. M. (1956). The relationship of hostility to authoritarianism. *The Journal of Abnormal and Social Psychology, 52*(3), 368.

Silber Mohamed, H. (2013). Can protests make Latinos "American"? Identity, immigration politics, and the 2006 marches. *American Politics Research, 41*(2), 298-327.

Sitter, K. C., & Curnew, A. H. (2016). The application of social media in social work community practice. *Social Work Education, 35*(3), 271-283.

Smith, E. (1993). Social identity and social emotions: toward new concepitualizations of prejudice. In D. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: interactive processes in group perception* (pp. 297-313). San Diego: Academic Press.

Smith, E. R., & Mackie, D. M. (2015). Dynamics of group-based emotions: insights from intergroup emotions theory. *Emotion Review, 7*(4), 349-354.

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can emotions be truly group level? Evidence regarding four conceptual criteria. *Journal of Personality and Social Psychology, 93*(3), 431.

Spata, C. (2015, October 29). Intern loses job after wildly racist tweet goes viral. Retrieved from http://www.complex.com/pop-culture/2015/10/intern-loses-job-racist-tweet-cotton-n-word

Stephan, W., & Stephan, C. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23-45). Hove, UK: Psychology Press.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of social issues, 41*(3), 157-175.

Stephan, W. G., Ybarra, O., & Bachman, G. (1999). Prejudice toward immigrants. *Journal of Applied Social Psychology, 29*(11), 2221-2237.

Stewart, J. (2012). *Fiction over facts: How competing narrative forms explain policy in a new immigration destination.* Paper presented at the Sociological Forum.

Stewart, T. L., Latu, I. M., Branscombe, N. R., & Denney, H. T. (2010). Yes we can! Prejudice reduction through seeing (inequality) and believing (in social change). *Psychological Science*.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior, 7*(3), 321-326.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence, 23*(04), 687-719.

Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association, 6*(2-3), 255-294.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations, 33*(47), 74.

Tajfel, H., & Turner, J. J. (1986). The social identity theory of intergroup behavior In S. Worchel & G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago, IL: Nelson.

Tam, T., Hewstone, M., Cairns, E., Tausch, N., Maio, G., & Kenworthy, J. (2007). The impact of intergroup emotions on forgiveness in Northern Ireland. *Group Processes & Intergroup Relations, 10*(1), 119-136.

Tape, T. G. (2001). Interpreting Diagnostic Tests. Retrieved from http://gim.unmc.edu/dxtests/Default.htm

Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British journal of social psychology, 48*(3), 427-446.

The Nexus fund. (2019). Verbal warning: Addressing Hate Speech and Dangerous Speech in the US . Retrieved from https://www.portlandoregon.gov/civic/article/677494

The Personal Responsibility and Work Opportunity Reconciliation Act, H.R.3734 P.L. 104-193 C.F.R. (1996).

The US Immigration and Nationality Act, Pub. L. No. 89-236, 27 Stat. 911 C.F.R. (1965).

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1).

Transue, J. E. (2007). Identity salience, identity acceptance, and racial policy attitudes: American national identity as a uniting force. *American Journal of Political Science, 51*(1), 78-91.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*: Basil Blackwell.

Turner, J. C., & Reynolds, K. J. (2011). Self-categorization theory. *Handbook of theories in social psychology, 2*, 399-417.

Tyler, I., & Marciniak, K. (2013). Immigrant protest: an introduction. *Citizenship studies, 17*(2), 143-156.

USDHS. (2014). (January 2014 ed.). Enforcement Integrated Database (EID): US Department of Homeland Security.

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American Behavioral Scientist, 57*(7), 920-942.

Valenzuela, S., Park, N., & Kee, K. F. (2009). Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. *Journal of Computer‐Mediated Communication, 14*(4), 875-901.

Van Assche, J., Roets, A., Dhont, K., & Van Hiel, A. (2014). Diversity and out-group attitudes in the Netherlands: The role of authoritarianism and social threat in the neighbourhood. *Journal of Ethnic and Migration Studies, 40*(9), 1414-1430.

Van de Vyver, J., Houston, D. M., Abrams, D., & Vasiljevic, M. (2016). Boosting belligerence: How the July 7, 2005, London bombings affected liberals' moral foundations and prejudice. *Psychological Science, 27*(2), 169-177.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107*.

Verkuyten, M., & De Wolf, A. (2007). The development of in-group favoritism: between social reality and group identity. *Developmental psychology, 43*(4), 901.

Verkuyten, M., & Hagendoorn, L. (1998). Prejudice and self-categorization: The variable role of authoritarianism and in-group stereotypes. *Personality and Social Psychology Bulletin, 24*(1), 99-110.

Verkuyten, M., & Martinovic, B. (2015). Majority member's recognition and protest against discrimination of immigrants: the role of power threat, deprovincialization and common national identity. *Social Justice Research, 28*(3), 257-273.

Viruell-Fuentes, E. A., Miranda, P. Y., & Abdulrahim, S. (2012). More than culture: structural racism, intersectionality theory, and immigrant health. *Social science & medicine, 75*(12), 2099-2106.

Wahl, O. F. (1997). *Media madness: Public images of mental illness*. New Jersey:: Rutgers University Press.

Wallace, S. J., Zepeda‐Millán, C., & Jones‐Correa, M. (2014). Spatial and temporal proximity: examining the effects of protests on political attitudes. *American Journal of Political Science, 58*(2), 433-448.

Wang, D. J., & Piazza, A. (2016). The use of disruptive tactics in protest as a trade-off: the role of social movement claims. *Social Forces, 94*(4), 1675-1710.

Waseem, Z., & Hovy, D. (2016). *Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter.* Paper presented at the SRW@ HLT-NAACL.

Wei, K., Jacobson López, D., & Wu, S. (2019). The Role of Language in Anti-Immigrant Prejudice: What Can We Learn from Immigrants' Historical Experiences? *Social Sciences, 8*(3), 93.

Wen, X., & Lin, Y. (2016). *Sensing Distress Following A Terrorist Event* Paper presented at the Social, Cultural, and Behavioral Modeling: 9th International Conference, SBP-BRiMS 2016, Washington, DC, USA, June 28-July 1, 2016, Proceedings.

Whitehead, J. W., & Aden, S. H. (2001). Forfeiting enduring freedom for homeland security: A constitutional analysis of the USA Patriot Act and the Justice Department's anti-terrorism initiatives. *Am. UL Rev., 51*, 1081.

Whitley Jr, B. E. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology, 77*(1), 126.

Williams, M. L., & Burnap, P. (2015). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology, 56*(2), 211-238.

Wodak, R., & Reisigl, M. (2001). *The Semiotics of Racism. Approaches in Critical Discourse Analysis*: Passagen Verlag.

Wolfe, S. E., Pyrooz, D. C., & Spohn, C. C. (2011). Unraveling the effect of offender citizenship status on federal sentencing outcomes. *Social Science Research, 40*(1), 349-362.

Wondracek, G., Holz, T., Kirda, E., & Kruegel, C. (2010). *A practical attack to de-anonymize social network users.* Paper presented at the Security and Privacy (SP), 2010 IEEE Symposium on.

Wong, K. C. (2006). The USA Patriot Act: A Policy of Alienation. *Mich. J. Race & L., 12*, 161.

Wu, E. D. (2014). *The color of success: Asian Americans and the origins of the model minority*. Princeton, New Jersey: Princeton University Press.

Yosso, T., Smith, W., Ceja, M., & Solórzano, D. (2009). Critical race theory, racial microaggressions, and campus racial climate for Latina/o undergraduates. *Harvard Educational Review, 79*(4), 659-691.

Young, E. (2014). *Alien Nation: Chinese Migration in the Americas from the Coolie Era through World War II*. Chapel Hill, NC: UNC Press Books.

Yzerbyt, V., Dumont, M., Wigboldus, D., & Gordijn, E. (2003). I feel for us: The impact of categorization and identification on emotions and action tendencies. *British journal of social psychology, 42*(4), 533-549.

Zakrisson, I. (2005). Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality and Individual Differences, 39*(5), 863-872.

Zárate, M. A., & Quezada, S. A. (2012). Future directions in research regarding attitudes toward immigrants. *Analyses of Social Issues and Public Policy, 12*(1), 160-166.

Zhang, W., Johnson, T. J., Seltzer, T., & Bichard, S. L. (2010). The revolution will be networked: The influence of social networking sites on political attitudes and behavior. *Social Science Computer Review, 28*(1), 75-92.

Zitek, E. M., & Hebl, M. R. (2007). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology, 43*(6), 867-876. doi:10.1016/j.jesp.2006.10.010

Zong, J., & Batalova, J. (2015). *Mexican Immigrants in the United States*. Retrieved from http://www.migrationpolicy.org/article/mexican-immigrants-united-states

APPENDIX A

Table 5.1 Protest locations for “Day Without Immigrants”

Protest city (State)	News sources about protest cities
Atlanta (Georgia)	http://www.ajc.com/news/national-govt--politics/day-without-immigrants-brings-parts-atlanta-area-halt/IZ3jWQm1QspAyMpfan6gHN/
Albuquerque (New Mexico)	http://www.kob.com/albuquerque-news/some-new-mexico-businesses-join-a-day-without-immigrants/4402332/
Dallas (Texas)	https://www.dallasnews.com/news/immigration/2017/02/16/nearly-50-dallas-area-businessesclose-day-without-immigrants-strike
Austin (Texas)	http://www.mystatesman.com/news/day-without-immigrants-call-embrace-austin-workers-families/ESsVjvgrZ8r4O7jZMSs5xL/
Denton (Texas)	http://www.thedentonite.com/blog/denton-day-without-immigrants
Fort Worth (Texas)	http://dfw.cbslocal.com/2017/02/16/day-without-immigrants-boycott-having-an-impact-in-north-texas/
Detroit (Michigan)	https://www.freep.com/story/news/local/michigan/detroit/2017/02/16/rally-forms-detroit-day-without-immigrants/97986924/
Aurora (Illinois)	https://www.cnbc.com/2017/02/16/mcdonalds-chains-shutter-in-support-of-a-day-without-immigrants-protest.html

Pittsburgh (Pennsylvania)	http://www.pennlive.com/news/2017/02/day_without_immigrants_protest.html
Philadelphia (Pennsylvania)	https://www.nbcphiladelphia.com/news/local/Philadelphia-Day-Without-Immigrants-Protest-Planned-413873993.html
Phoenix (Arizona)	http://www.phoenixnewtimes.com/restaurants/phoenix-restaurants-participating-in-day-without-immigrants-protest-today-9090123
Riverside (California)	http://abc7.com/education/riverside-teachers-put-on-paid-leave-over-immigrant-protest-comments/1766602/
San Francisco (California)	http://www.sfgate.com/bayarea/article/I-am-ashamed-Store-owner-who-voted-for-10938441.php
New York (New York)	https://www.nytimes.com/2017/02/16/nyregion/day-without-immigrants-boycott-trump-policy.html
Washington DC	http://www.cnn.com/2017/02/15/politics/washington-dc-day-without-immigrants/index.html
Boston (Massachusetts)	https://www.bostonglobe.com/lifestyle/food-dining/2017/02/16/what-closed-boston-area-for-day-without-immigrants-protests/61js7eG4U9SusMeUFSOgEP/story.html
Portland (Oregon)	http://www.kgw.com/news/politics/portland-businesses-expected-to-participate-in-day-without-immigrants/408837243

Table 5.2 Protest locations for "No ban no wall" protest

Protest city (State)	News sources about protest cities
New York (New York)	https://nypost.com/2017/01/29/massive-protests-erupt-at-airports-across-us-after-trumps-immigrant-ban/
Atlanta (Georgia)	https://www.wsbtv.com/news/local/atlanta/protests-against-trump-travel-ban-to-gather-at-hartsfield-jackson/488947556
San Francisco (California)	https://www.elpasotimes.com/story/news/2017/01/28/1000-flood-sfo-protest-immigration-ban/97200022/
Los Angeles (California)	https://www.usatoday.com/story/news/nation/2017/01/29/homeland-security-judges-stay-has-little-impact-travel-ban/97211720/
Sacramento (California)	https://www.sacbee.com/news/local/article129482184.html
Mountain View (California)	https://www.nbcnews.com/tech/tech-news/google-employees-rally-against-president-trump-s-immigration-ban-n714466
Chicago (Illinois)	http://www.timesofisrael.com/history-in-mind-jews-across-us-join-airport-protests-of-refugee-ban/
Detroit (Michigan)	https://detroit.cbslocal.com/2017/01/29/protests-against-muslim-ban-taking-place-at-metro-airport/
Indianapolis (Indiana)	http://www.indystar.com/story/news/2017/01/29/hundreds-gather-protest-immigration-order-indy-airport/97224888/
Portland (Maine)	https://www.pressherald.com/2017/01/29/anti-trump-rallies-are-popping-up-around-maine-today/

Boston (Massachusetts)	https://www.npr.org/sections/thetwo-way/2017/01/29/512250469/photos-thousands-protest-at-airports-nationwide-against-trumps-immigration-order
Cleveland (Ohio)	https://www.wkyc.com/article/news/protesters-crowd-cleveland-hopkins-airport-for-demonstration/394316164
Columbus (Ohio)	https://www.cnn.com/2017/01/30/politics/travel-ban-protests-immigration/
Orlando (Florida)	http://floridapolitics.com/archives/231195-1000-gather-orlando-airport-protest-donald-trumps-immigrant-ban
Philadelphia (Pennsylvania)	https://www.delawareonline.com/story/news/local/2017/01/29/delaware-immigration-trump-protest/97215226/
Portland (Oregon)	https://katu.com/news/local/protest-disrupts-traffic-at-portland-international-airport
Raleigh–Durham (North Carolina)	https://www.indyweek.com/news/archives/2017/01/29/1500-protesters-converge-on-rdu-to-protest-trumps-muslim-ban
Seattle (Washington)	https://patch.com/washington/seattle/more-2-000-protest-trump-executive-order-sea-tac-airport
Austin (Texas)	https://www.statesman.com/news/hundreds-protest-muslimban-austin-airport/PZe2fObvp0Scpj1DNQmTUP/
Dallas (Texas)	https://www.npr.org/sections/thetwo-way/2017/01/29/512250469/photos-thousands-protest-at-airports-nationwide-against-trumps-immigration-order

Houston (Texas)	https://www.chron.com/houston/article/Protests-planned-against-Muslim-ban-across-10892379.php
Washington DC	https://www.cnn.com/2017/01/29/politics/immigration-crisis-what-we-know/

APPENDIX B

Manual inspection criteria	Included words (n = 161)
misspelled	immigrats, immigran, immigrat, immigr, immegrants, inmigrant, immgrnts, immigrts, immigrantwho, immi, imms, imigrant, imagrants, immigrati,'immigrates, imm, inmigrants,immagrants, immigs, immigant, immigra, immig, immigrantscall, immagrant, imigrants,
synonyms	foreigners, migrant, alliens, undocumenteds, emigrants,refugee, refug, expatriate, undocu, aliens, unau, migrants, alien, unauthorised, lprs, undocumented, asylee, newcomers, undocume, undoc, refuges, settlers,
figurative language	illegalls, illgls, illegalaliens, illegals, legals, illigals, undesirables, rapists, dacamented, infiltrators, terrorists, extremists, rapeugees, schooljihadi, criminals, murderers, jihadis, terroists, moslems, jihadists, rapefugees, rapefugee, lawbreakers, invaders, scroungers, invader, dreamer, freeloaders, invasions, wetbacks, dreamers,
group related to immigrants	filipinos, musli, salvadorans, sibrian, muslim, palestinians, expat, mexicans, cubans, hispanics, nigerians, citizen, afghans, refugees, latinxs, muslims, italians, israelis, arabs, ethiopian, citizens, hispanic, syrian, latino, residen, kenyans, descendants, citzens, iraqis, afghani, laborers, biafrans, expats, uzbek, honduran, indian, muslin, ctzns, deportees, iranians, latinx, pakistanis, pakistani, latinos, syrians, iranian, citizen, somalis, somali,
action related to immigrants	immigration, immagration, expelling, illegally, illeg, immigrantion, deportations, assimilated, immigrations, repatriation, emigration, assimilation,

	deportation, muslins, borders, migration, deporting, farmworkers, illegal, imigration, unassimilated, illega, iligal, illigal, roundups, nationalities, llegal, ilegal, assimilating, imagration, migrates, influx

APPENDIX C

Coding Instructions for Classifying Prejudiced Speech against Immigrants

We appreciate your help with this research project. **The goal of this project** is to validate a method that measures prejudiced speech against immigrants. **Your task** is to decide whether a Twitter post (hereafter, tweet) expresses prejudice against immigrants. **In the following section**, we will provide

1. Definition of immigrants
2. Indicators, examples, and rationales for coding a tweet as about immigrants
3. Definition of prejudiced speech against immigrants,
4. Indicators, examples, and rationales for coding a tweet as prejudiced speech against immigrants, and
5. Coding procedure.

1. Definition of immigrant

Immigrants are ***foreign-born persons*** who have migrated to a country on ***a long-term*** basis, which includes those who are, or are in the process of becoming, naturalized citizens or lawful permanent residents; those who are refugees or asylum-seekers; and those who reside in the country without government permission.

2. Indicators, examples and rationales for coding a tweet as about immigrants

A tweet is considered as about immigrants if it:

1) **discusses immigrants**

- Examples: Hunters claimed immigrants attacked them near the US-Mexico border. Investigators say the hunters shot each other.
- Rational: This tweet is a about immigrant because it discussed “immigrants”.

- Examples: Raids across the US leave immigrant communities and activists on high alert
- Rational: This tweet is a about immigrant because it discusses “immigrant communities”.

2) **discusses groups of immigrants.**

- Examples: If any of these refugees commit a crime or an act of terror. The blood is on ur hands leftists. Ur grimy nasty indecent little hands.
- Rational: This tweet is a about immigrants because it discusses “refugees”.

- Examples: Since those German immigrants from the 1800's are taking our jobs of making shitty beer. #boycottbudwiser
- Rational: This tweet is a about immigrants because it discusses “German immigrants”.

- Examples: Miami-Dade Mayor Carlos Gimenez should have protected illegal immigrants, not bowed down 2 Trump's sanctuary order
- Rational: This tweet is a about immigrants because it discusses “illegal immigrants”.

3. **Definition of prejudiced speech against immigrants**

   By prejudiced speech against immigrants, we mean a tweet that convey antipathy or hostility directed toward immigrants.

4. **Indicators, examples, and rationales for coding a statement as prejudiced speech against immigrants**

   A tweet is considered as prejudiced speech against immigrants if it:

   1) **insults, belittles, curses, or stigmatizes immigrants**

      - Example: "All You Fuckin Immigrants Fucked".
      - Rational: This statement expresses prejudiced speech against immigrants because it uses a swear word "Fuckin" to curse immigrants.

      - Example: "We must not "normalize" Mexican Invaders, we must call them out. Do not use "immigrant" avoid it, deport all of them"

      - Rational: This statement expresses prejudiced speech against immigrants because it uses words "Mexican Invaders" to stigmatize Mexican immigrants.

      - Example: "Screw the immigrants who did it legally. Screw the Americans who face lower wages, and lose jobs due to cheap labor."

- Rational: This statement expresses prejudiced speech against immigrants because it uses “screw” to curse immigrants.

2) **criticizes immigrants for damaging the economy, such as by stealing jobs, abusing taxpayer’s money, and depleting welfare resources**
    - Example: “Check out these jerk immigrants stealing your jobs.”
    - Rational: This statement expresses prejudice against immigrants because it criticizes immigrants for “stealing jobs”.

    - Example: “ANN COULTER ON IMMIGRANTS LEECHING USA’S WELFARE. ANCHOR BABY SHOULDN’T B LEGAL.”

    - Rational: This statement expresses prejudice against immigrants because it criticizes immigrants for depleting welfare resources, indicated by “IMMIGRANTS LEECHING USA'S WELFARE” .

    - Example: “Strange how MOST Muslim Immigrants are living on OUR Tax Dollars, And the tax dollars of Most countries they immigrate to”
        - Rational: This statement expresses prejudice against immigrants because it criticizes “Muslim immigrants” for “living on OUR Tax Dollars”.
3) **criticizes immigrants for damaging the culture, such as by lacking knowledge about English skills, lacking intention to assimilate and integrate to the host country, and contaminating religious purity**

- Example: "We need to stop assuming that immigrants will integrate."
- Rational: This statement expresses prejudice against immigrants because it assumes immigrants having no intention to integrate to our country.

- Example: "I'm SO SICK n TIRED of #AntiWhite immigrants using our hospitality to try and DESTROY White Christian America"
- Rational: This statement expresses prejudice against immigrants because it assumes immigrants have the intention to "DESTROY White Christian America".

4) **criticizes immigrants for damaging public safety such as by posing threats to security and increasing crime rates**
    - Examples: "Somali Immigrant Trained with Terrorist Group, Plotted Domestic Attack Imagine that....WHY EE NEED THE #TravelBan "
    - Rational: This statement expresses prejudice against immigrants because it associates Somali immigrants with terrorists.

    - Example: "Yay let's let them in! We should actually get rid of security all together. Love all immigrants. (Do I sound crazy to y'all too?) #MAGA"
    - Rational: This statement expresses prejudice against immigrants because it criticizes immigrants for posing threats to security.

- Example: “Libtards don't get it Crime rate in Sweden has skyrocketed because of Immigrants”
- Rational: This statement expresses prejudice against immigrants because it assumes that immigrants increase crime rates.

5) **opposes immigration such as by supporting strict immigration laws that exclude immigrants or supporting distorted views about immigrants in news**
    - Example: “PODUS keep doing what U r doing no more immigrants 4 now Nationality Act”
    - Rationale: This statement expresses prejudice against immigrants because it against immigrants, “no more immigrants” and it refers to “Nationality Act”, which suggests restricting immigration based on country of origins.

    - Example: “#DayWithoutImmigrants is great, DECADES without immigrants would be even better, they should all go back to their home”
    - Rationale: This statement expresses prejudice against immigrants because it is against immigration and intends to exclude immigrants.

5. Coding procedure.

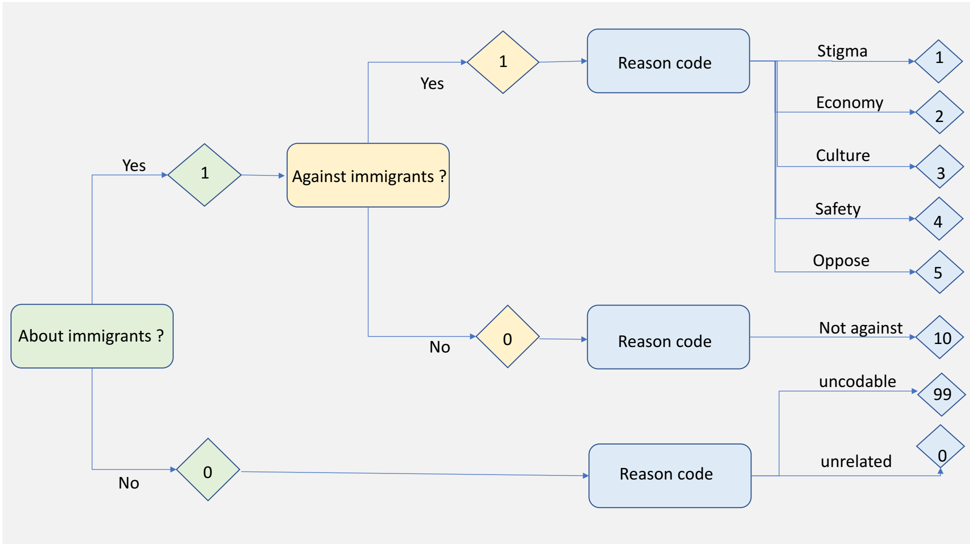

Figure 1 is a work flow for coding prejudice against immigrants.

First, you need to use *definition of immigrants* to decide whether a statement is *about immigrants*.

- If it is about immigrant, code 1 and proceed to code whether it is against immigrants.
- If it is not about immigrants, code 0 and provide your reasons under reason code:
  - Uncodable: 99
  - unrelated: 0.

Second, you need to use *definition and indicators of prejudice against immigrants* to decide whether a statement is against immigrants.

- If it is against immigrants, code 1 and provide your reasons under reason code:
  - Stigma: 1;
  - Economy: 2;
  - Culture: 3;

- Safety: 4;
- Oppose: 5

- If it is not against immigrants, code 0 and provide your reasons under reason code:
  - Not against: 10.

Table 1. Description for Reason code

Reason	Reason code	Description
Stigma	1	This statement insults, belittles, curses, or stigmatizes immigrants
Economy	2	This statement criticizes immigrants for damaging the economy, such as by stealing jobs, abusing taxpayer's money, and depleting welfare resources.
Culture	3	This statement criticizes immigrants for damaging the culture, such as by lacking knowledge about English skills, lacking intention to assimilate and integrate to the host country, and contaminating religious purity.
Safety	4	This statement criticizes immigrants for damaging public safety such as by posing threats to security and increasing crime rates
Oppose	5	This statement opposes immigration such as by supporting strict immigration laws that exclude immigrants or supporting distorted views about immigrants in news
Not Against	10	This statement is about immigrants, but it does not express prejudice against immigrants.
Uncodable	99	I don't understand the statement.
Unrelated	0	This statement is not related to immigrants.

Table 2. Coding Examples

Text	**About immigrants?**	**Against immigrants?**	**Reason code**
All You Fuckin Immigrants Fucked.	1	1	1
PODUS keep doing what U r doing no more immigrants 4 now Nationality Act	1	1	2
Check out these jerk immigrants stealing your jobs	1	1	3
I'm SO SICK n TIRED of #AntiWhite immigrants using our hospitality to try and DESTROY White Christian America	1	1	4
Somali Immigrant Trained with Terrorist Group, Plotted Domestic Attack Imagine that....WHY EE NEED THE #TravelBan	1	1	5
Your father is responsible for two senior citizens with green cards being illegally detained at O'Hare	1	0	6
I saw it on TV tonight. Another Liberal jab + market to cater to Latinos. Junk beer. Thks 4 the heads up on. BB. I'll pass	0	0	0
Finding people who are not part of WeThePeople is a health exam looking 4nonAmericanImmigrants/refugees who fraud vote	1	0	99