Effects of Dehumanization and Disgust-Eliciting Language on Attitudes Toward Immigration: A Sentiment Analysis of Twitter Data

Katherine S. Wahrer
University at Albany, State University of New York

Cynthia J. Najdowski
University at Albany, State University of New York, cnajdowski@albany.edu

John V. Passarelli

Follow this and additional works at: https://scholarsarchive.library.albany.edu/psychology_fac_scholar

Part of the American Politics Commons, Communication Technology and New Media Commons, Demography, Population, and Ecology Commons, Gender, Race, Sexuality, and Ethnicity in Communication Commons, Immigration Law Commons, Interpersonal and Small Group Communication Commons, Law and Psychology Commons, Mass Communication Commons, Social Influence and Political Communication Commons, Social Justice Commons, Social Media Commons, Social Psychology Commons, Social Psychology and Interaction Commons, and the Speech and Rhetorical Studies Commons

Recommended Citation
Wahrer, Katherine S.; Najdowski, Cynthia J.; and Passarelli, John V., "Effects of Dehumanization and Disgust-Eliciting Language on Attitudes Toward Immigration: A Sentiment Analysis of Twitter Data" (2024). Psychology Faculty Scholarship. 48.
https://scholarsarchive.library.albany.edu/psychology_fac_scholar/48

Rights Statement
This Article is brought to you for free and open access by the Psychology at Scholars Archive. It has been accepted for inclusion in Psychology Faculty Scholarship by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Effects of Dehumanization and Disgust-Eliciting Language on Attitudes Toward Immigration:
A Sentiment Analysis of Twitter Data

Katherine S. Wahrer, M.A. a
Cynthia J. Najdowski, Ph.D. b
John V. Passarelli, B.A. c

a University at Albany, State University at New York, School of Criminal Justice, Draper Hall
Room 219, 135 Western Avenue, Albany, NY 12222

b University at Albany, State University of New York, Department of Psychology, Social
Sciences Building, Room 399, 1400 Washington Avenue, Albany, NY 12222

c Independent Researcher

Word Count for manuscript with figures, tables and references: 13,239
Abstract

Attitudes towards immigration have been shown to be driven by dehumanization and disgust. The more people dehumanize immigrants and the more disgusted they feel, the more negative attitudes they tend to have toward immigrants. However, little is known about how exposure to social media content that links dehumanization, disgust, and immigration influences users’ attitudes on this issue. This is important to consider because the majority of adults in the United States are on social media. We used Twitter data, machine learning, and sentiment analysis to investigate whether exposure to dehumanizing or disgust-eliciting tweets about immigration impacts users’ own sentiment toward immigration over time. Our results were in some ways consistent and in other ways inconsistent with prior literature. They showed that either dehumanizing or disgust-eliciting language appears in 66% of our sample of tweets pertaining to immigration. Unexpectedly, however, exposure to both kinds of language in tweets about immigration related to small increases in positive sentiment about immigration over time. There was evidence of Granger-causality only for dehumanizing language, however, and only when controlling for the political affiliation of the communicator. These findings indicate that social media exposure may influence public perceptions of immigrants and immigration issues in unexpected ways.
1.1 Background

When announcing the end of the longest shutdown of the U.S. government in history on January 25, 2019, former President Donald J. Trump identified differing opinions about immigrants and immigration as the central point of political debate that contributed to the shutdown. He stated:

The sheer volume of illegal immigration has overwhelmed federal authorities and stretched our immigration system beyond the breaking point. Nearly 50 migrants a day are being referred for medical assistance—they are very, very sick—making this a health crisis as well. It’s a very big health crisis. (Trump, 2019, para. 26)

Trump’s reference to Mexican immigrants as “volume” is an example of the type of language that has been used throughout history to de-identify “others” and mark them as less than human (Markel & Stern, 2002). The comment that migrants are “very, very sick” was likely made in an effort to elicit a disgust reaction from his audience. This type of dehumanizing and disgust-eliciting language is frequently used in reference to immigrants not only in speeches such as Trump’s, but also in other forms of communication such as print media (O’Brien, 2003). Prior research has found that the presentation of such metaphors in traditional media sources like newspapers or magazines can influence people to dehumanize immigrants, report higher levels of disgust, and feel more negatively toward them (Buckels & Trapnell, 2013; Hodson & Costello, 2007; Utych, 2018). Social media platforms have become a common venue for sharing views and attitudes, but it remains unclear how frequently dehumanizing and disgust-evoking metaphors are used in the context of social media, or whether exposure to such language via social media platforms impacts users’ attitudes toward immigrants. The current study addresses
these gaps in the literature by analyzing data from Twitter and determining how exposure to dehumanizing or disgust-eliciting language impacts users’ sentiment toward immigrants.

These are important issues to explore. One concern is that the use of such language in propaganda may help promote certain policies desired by political elites, as suggested by the earlier quotation. For example, the “ Victims of Immigrant Crime Engagement Office” (VOICE) was created in 2017 during a time when immigration and the push for funding a wall on the U.S. border with Mexico were in the media spotlight. Their mission was to compile and publicize weekly reports of crimes committed by undocumented immigrants, thus intending to “fault or scapegoat migrants” (Pérez Miles, 2017, p. 8). This is notable as dehumanizing or disgust-eliciting language describing immigrants as “criminal,” “rapists,” a “flood,” or a “swarm” can potentially illicit anxiety and spur political action (Brader et al., 2008). As Esses et al. (2013) described, “uncertainty can be used to media and political advantage, allowing the transformation of relatively mundane episodes into newsworthy events that can be sold to the public and can serve as support for relatively extreme political platforms” (p. 519). In fact, research has shown that people who perceive immigrants as threatening are significantly more likely than others to have negative attitudes toward immigrants and to support stringent anti-immigration policy (Lewis et al., 2013).

Another concern is that there can be serious ramifications for society when dehumanizing and disgust-eliciting language is used in everyday communication (Utych, 2018). Even though the main function of the media is to spread information, it also can be manipulated or distorted to portray others in a negative light (Pérez Miles, 2017). If dehumanizing and disgust-eliciting metaphors and stereotypes continue to be used to depict immigrants and immigration in public discourse, then society may adopt this negative connotation as the norm (Hodson et al., 2014).
This could lead immigrants to be excluded from the protections held for those in the moral community. In turn, exclusion of immigrants from the moral community could potentially contribute to aggressive and punitive responses to them, hate crimes, or, even worse, genocide (Müller & Schwarz, 2020; Opotow, 1990; Stanton, 2016). As evidence of this, Bandura et al. (1975) found that participants who were told they could punish people were most likely to punish those who were characterized in dehumanizing rather than neutral or humanizing terms. Further, it was the ability for participants to deny human nature to others that acted as the most potent disinhibitor of punishment. Notably, Kelman (1973) identified the denial of “identity” to others as a precursor for sanctioned mass violence. The use of metaphors to shape public opinion has been a precursor to mass violence for centuries because cultural violence depicted through day-to-day language can legitimize direct violence (Galtung, 1990). For instance, such dehumanization can be seen in the metaphorical depictions of Jews as vermin that were common during the Holocaust, and of Hutus as cockroaches by Tutsis during the Rwandan Genocide (Haslam, 2006; O’Brien 2003; Staub, 1999; Utych, 2018). Although Trump’s public dehumanizing and disgust-eliciting comments about immigration may seem relatively benign compared with genocides, it is important to acknowledge how such language may be a precursor to harm. Indeed, Trump has endorsed violence against migrants both publicly and privately, including suggesting that soldiers should shoot migrants (Shear & Davis, 2019). Moreover, his anti-Muslim social media posts have been linked to an increase in anti-Muslim sentiment within his follower’s posts as well as anti-Muslim hate crimes (Müller & Schwarz, 2023); his tweets referring to the “Chinese virus” at the outset of the COVID-19 pandemic also have been connected to a rise in anti-Asian hate crimes (Cao et al., 2023).
In light of the potential negative and serious political and societal impacts that may result from language that dehumanizes or elicits disgust toward immigrants, we review the history of such language on the issue, the evolutionary roots of dehumanization and disgust in relation to their contemporary manifestations, and existing evidence that exposure to such language affects one’s attitudes toward immigration. From this foundation, we extrapolate to consider how and to what effect dehumanizing and disgust-eliciting language is used in the novel context of social media. Throughout we refer to migrants/migration and immigrants/immigration without making distinctions regarding legality. Although attitudes toward illegal immigrants tend to be more negative than attitudes toward legal immigrants (see Esses, 2021, for review), contemporary media discourse regarding immigration focuses heavily on the U.S.-Mexico border, and Mexican immigration is often conflated with illegal immigration (Dick, 2011; Kinefuchi & Cruz, 2015). We consider this point further in the discussion section.

1.1.1 Dehumanization and Disgust in Immigration Rhetoric

Theory on dehumanization posits that there are certain characteristics that we may deny to “others” (Haslam, 2006). This can be exhibited in a variety of ways, including in the language used to refer to people or even through discrimination. Typically, dehumanization occurs by denying others either (a) uniquely human attributes or (b) human nature qualities (Haslam, 2006). To deny someone uniquely human attributes is to see them as childish, aggressive, impulsive, and lacking in self-control (Haslam, 2006). This animalistic dehumanization (Haslam, 2006) or infrahumanization (Leyens et al., 2000) is conveyed, for instance, in historical depictions of Black people as “apes” or “monkeys” (Kendi, 2016). Another example is that Jews and Bosnians were thought of as “vermin” during their respective killings in the Holocaust and Srebrenica genocide (Chalk & Jonassohn, 1990; Kelman, 1973). The second form of
dehumanization is referred to as mechanistic dehumanization, through which people are denied the qualities of human nature such as personal depth, emotional responsiveness, and agency; thus, one would think of someone as an object, cold, lacking emotions, and indifferent to most things. For example, certain people who emigrated to the U.S. in the early 1920s were referred to as “the dumping of Europe’s human refuse at our doors” (Weber, 1892, p. 424). As these examples demonstrate, dehumanizing metaphors have been used to describe immigrants across geographical space and throughout time.

These metaphors persist, as one study of major U.S. magazines (i.e., *Time, Newsweek*) conveyed that themes portraying immigrants as invaders appeared consistently from 1965 to 1999 (Chavez, 2001). This sort of dehumanizing language also was seen in a study of Australian print media about asylum seekers from 2001 to 2002 (Klocker & Dunn, 2003). The results showed that 76% of print media portrayed asylum seekers as “terrorists” or “criminals.” Similarly, Cisneros (2008) found that U.S. television in 2005, largely portrayed immigrants as a dangerous pollutant that needed to be controlled to prevent contamination. Consider even the frequently used names for non-citizen immigrants: “illegals,” “aliens,” and even “illegal aliens”; all are dehumanizing. Indeed, the U.S. Biden-Harris administration acknowledged this in their proposed immigration reform bill, describing the term “alien” as a “dehumanizing slur” (Acevedo, 2021, p. 1).

In addition to dehumanizing metaphors, disgust metaphors have been used to describe immigrants for scores of years (O’Brien, 2003; Rozin, Haidt & McCauley, 2008). In fact, anti-immigration rhetoric has been intertwined with medical language concerning diseases since the inception of the U.S. (Markel & Stern, 2002; O’Brien, 2003). In the late 19th century, anti-immigration rhetoric tended to have a health focus because immigrants brought new and
unknown diseases to the shores of the U.S. (Markel & Stern, 2002). For instance, the 1882 Chinese Exclusion Act was enacted largely due to southern China being a part of the “hookworm belt” during the time (Markel & Stern, 2002; O’Brien, 2003). Therefore, strict stances against immigration were taken partially due to the desire to avoid the spread of diseases. However, the association between disgust and immigration remains prevalent to this day. This is evident in the quotation from former President Trump noted earlier, in which he referred to migrants as “very, very sick” and described the influx of migrants at the border as a “health crisis” (Trump, 2019). Further, as Marshall and Shapiro (2018, p. 774) state, “the current discourse in the U.S. surrounding unauthorized immigrants includes metaphors that readily activate thoughts of vermin (e.g., rodents).” While this dehumanizes immigrants by diminishing them to the subhuman category of rodents, it simultaneously elicits a disgust reaction. Indeed, the connection between dehumanization and disgust in relation to immigration is no coincidence, as discussed next.

1.1.2 The Evolutionary Connection: Dehumanization and Disgust

Disgust is an innate emotion that aids in the avoidance of potentially hazardous materials (Rozin et al., 1994; Rozin et al., 2008). But early humans also had to avoid dangerous unknown groups to survive. Similar to how humans developed taste aversion to avoid contaminated food, our behavioral immune systems may have developed disgust reactions to avoid potentially contaminated people (Rozin et al., 2008). This evolutionary drive to stay away from poisonous foods, diseases, or potentially contagious people is engrained in each of us (Aarøe et al. 2017; Faulkner et al., 2004; Navarrete & Fessler, 2006). In fact, neurological evidence supports the evolutionary connection between disgust and dehumanizing reactions to outgroups. For instance, Harris and Fiske (2006) examined the neurological activity of participants viewing marginalized
individuals (e.g., homeless people, drug addicts). They found that viewing these groups led to neural activation in the amygdala and insula, brain structures that are commonly associated with disgust. This is consistent with other research showing that all forms of disgust activate the amygdala, which is the area of the brain that helps associate things that should be avoided with fear (Herz, 2011). Furthermore, and of crucial importance, the participants in Harris and Fiske’s (2006) study had no neural activation in the medial prefrontal cortex, the area typically involved when humans engage in social cognition. This pattern of neurological activity suggests that the participants perceived marginalized individuals as less than human.

Rozin et al. (1997, p. 73) describe this innate connection between disgust and dehumanization by noting that “disgust reactions connote the sense that one is better, purer, and less offensive than the offending target . . . disgust serves as an outgroup marker.” Thus, visceral disgust reactions allow humans to biologically distinguish the “ingroup” from the “outgroup.” However, this evolutionarily based association ceases to have benefits when the behavioral immune system signals humans to avoid “others” who are unlikely to be carrying pathogens. Even so, research suggests that the connection between disgust and dehumanization does in fact influence day-to-day intergroup interactions. For example, Buckels and Trapnell (2013) found that participants who were randomly assigned to feel disgusted were more likely than other participants to associate outgroup names with animals and ingroup names with humans on an implicit association test (IAT). In other work, Hodson and Costello (2007) found that the more sensitive participants were to interpersonal disgust (e.g., not wanting to share other people’s clothes or drink from the same cups as others), the more they endorsed dehumanizing perceptions of immigrants. Of importance, interpersonal-disgust sensitivity also predicted negative attitudes toward immigrants and other foreign groups. Thus, the pairing of disease and
other dehumanizing terms such as worms or vermin with the image of “immigrants” helps to strengthen a pattern of behavioral avoidance of foreigners (Markel & Stern, 2002; O’Brien, 2003).

1.1.3 Effects of Dehumanization and Disgust in Traditional Media on Attitudes toward Immigration

Due to the social psychological and neurobiological evidence that links dehumanization and disgust, it is important to consider how the use of metaphors that evoke these reactions in everyday communications translate into attitudes toward immigrants and immigration. Research has shown that the combination of dehumanizing and disgust-eliciting language, such as the phrase “filthy pigs,” can influence others to have more negative attitudes toward immigrants (Hodson et al., 2014). For instance, Utych (2018) found that participants who were assigned to read text with disease metaphors designed to dehumanize immigrants reported more negative attitudes toward immigrants than participants in a comparison group who were not exposed to such language. Further, the participants in the dehumanization group indicated they felt significantly higher levels of disgust toward immigrants than the comparison group. In another study, participants were randomly assigned to read a news article that either did or did not include “vermin” metaphors (Marshall & Shapiro, 2018). This work revealed that those exposed to the dehumanizing and disgust-evoking metaphors reported feeling significantly higher levels of disgust than those not exposed to the “vermin” metaphors. Moreover, within the “vermin” condition, those with higher versus lower disgust sensitivity showed stronger disapproval of close contact with immigrants and supported more stringent immigration policies.

In the same vein, Esses et al. (2013) experimentally manipulated an “incidental” viewing of an editorial cartoon of an immigrant carrying a suitcase with labels of various infectious
diseases applied in one condition and no labels in the other. They then asked participants several questions about their memory of the cartoon to test whether the labels primed certain attitudes toward immigrants. Compared to participants who were not shown disease labels, those who were exposed to the incidental disease labels were significantly more likely to report thinking immigrants were spreaders of disease, to dehumanize and lack admiration for immigrants, and to have less favorable attitudes toward immigrants and immigration overall. Moreover, these effects manifested regardless of whether the exposed participants actually recounted seeing the labels. Other studies have found that the association between disgust and dehumanization is related to policy preferences such as support for deportation (Dalsklev & Rønningsdalen-Kunst, 2015). For example, participants from Norway read a newspaper clipping describing members of Roma, a marginalized ethnic group, as having either low hygiene standards or good hygiene. Those who read about low hygiene standards were significantly more likely than others to dehumanize Roma people and support their deportation (Dalsklev & Rønningsdalen-Kunst, 2015). Together, these studies suggest that the dehumanizing or disgust-eliciting metaphorical depiction of immigrants by the media can influence both attitudes about immigration and policy preferences.

1.1.4 The Importance of Social Media

Today, people are increasingly turning to social media to learn about their world (Greenwood et al., 2016) and share opinions. Shearer and Matsa (2018) estimated that about 68% of the U.S. population at least occasionally receives news from social media. At the time this study was conducted, one social media platform in particular, Twitter, had the highest rate of news-focused users, with at least 15% of U.S. people reporting using Twitter every day (Flores et al., 2017). Thus far, this source of public opinion data remains largely unstudied in the
ATTITUDES TOWARD IMMIGRATION

criminological literature, even though it has been shown to reflect public attitudes, sentiments, and even behavior (for reviews, see Flores, 2017; Müller & Schwarz, 2023). For example, attitudes expressed in tweets have been shown to have a robust correlation with those reported on public opinion surveys (O’Connor et al., 2010) and correspond with current influenza outbreaks (Culotta, 2010) and criminal activity (Wang et al. 2012), particularly hate crimes (Cao et al., 2023; Müller & Schwarz, 2023).

Chadwick (2013) describes social media as a “hybrid media system.” Rather than a landscape dominated by the voices of the powerful, everyone is invited to share their opinions. The ability for social media to provide a platform to all perspectives could have a variety of effects. One concern is related to echo chambers. People tend to seek out media and information that confirms their preexisting beliefs (Sears & Freedman, 1967), and social media users’ beliefs are positively reinforced through “likes” or “followers” that they attract (Lemieux & Regens, 2012). However, users also can choose to hide, unfollow, or block those that disagree with them. By doing so, users may create their own echo chambers which expose them mostly to ideas similar to their own. These “information bubbles” often radicalize rather than relativize users’ existing opinions (de Saint Laurent et al., 2020). This could negatively impact societal perceptions of immigrants, as people with unfavorable attitudes toward immigrants or immigration are unlikely to have their views challenged by others due to these chambers. Of relevance to the current study, users in echo chambers may be exposed to social media messages containing dehumanizing or disgust-eliciting language about immigration and, as a result, adopt increasingly negative attitudes on the issue over time.

Alternatively, social media can be a fertile ground for contentious exchanges between people holding opposing viewpoints (Klein, 2019). Thus, it could be used by anti-immigration
users as a space for exposing and indoctrinating those who hold more positive attitudes toward immigrants. de Saint Laurent et al. (2020) analyzed Twitter to understand how anti-immigration users operate on social media. They found that, in comparison to pro-immigration users, anti-immigration users were more active on Twitter and used more creative slogans and hashtags to stigmatize immigrants as dangerous and criminal. They concluded that anti-immigration users purposely aim to expand their community outside of their echo chamber and influence online interactions with the goal of negatively affecting immigrants and their prospects. Dehumanizing and disgust-eliciting language may be one tool employed for this purpose.

Regardless of whether social media users are communicating with others who share their attitudes toward immigration or oppose them, an additional concern is linguistic mimicry. Linguistic mimicry relates to how closely people match others’ word use or style in conversation (Moore & McFerran, 2011), similar to behavioral mimicry by which one person copies the behavior of another. Linguistic mimicry can act “as a ‘social glue’ that both reflects and creates bonds between people” (Moore & McFerran, 2017, p. 18). Notably, this occurs even when conversations are online, regardless of the relationship between the conversation holders (Niederhoffer & Pennebaker, 2002). Further, even if two people engaging in an online conversation dislike one another or disagree, they still tend to mimic each other’s language. Linguistic mimicry has been shown to change online communicators’ attitudes as a result of the associative pairing that it produces (e.g., Moore & McFerran, 2011). The question is whether people mimic each other’s immigration-relevant dehumanizing and disgust-eliciting language and metaphors online and come to increasingly dehumanize or have disgust reactions to immigrants or immigration over time, regardless of their initial attitude positions or whether such changes are consciously endorsed.
To summarize, little is understood about how dehumanizing and disgust-eliciting language impact attitudes on social media where people are able to interact with it, but it is clear at least that there are multiple possible paths by which the language may exert a negative influence. It is important to elucidate whether this happens given other evidence that social media users use dehumanizing and disgust-eliciting language in their posts related to immigration to at least some extent. For instance, Dhuman Giron and Vargas (2020) found that messages on Twitter conveyed components of the Latino threat narrative (i.e., that Latinx are taking over the continental U.S.), thereby vocalizing the dehumanizing narrative that this population is “dangerous, crime-ridden, and undesirable” (p. 115). Musolf’s (2015) analysis of anti-immigration blogs found that users frequently referred to immigrants as leeches, locusts, rats, vermin, plague, germs, and contaminated, thus “denouncing them as being both metaphorically and literally dirty” (Musolf, 2015, p. 50). Neither the prevalence of such language on social media nor its effects on other users is known, however. This is crucial to understand because it could contribute to anything from support for deportation to hate speech or even violence (Cao et al., 2023; Dalsklev & Rønningsdalen-Kunst, 2015; Hankes & Amend, 2018; Klein, 2019; Müller & Schwarz, 2023).

1.2 The Current Study

Dehumanization and disgust-eliciting rhetoric about immigration has not been studied in the context of social media, even though increasingly people use social media to learn about their world. Research suggests that dehumanizing and disgust-eliciting language influence those exposed via traditional print media to adopt more negative attitudes towards immigration (Dalsklev & Rønningsdalen-Kunst, 2015; Esses et al., 2013; Hodson et al., 2014; Marshall & Shapiro, 2018; Utych, 2018), but whether these effects generalize to the social media context
remains unknown. The impact of dehumanizing and disgust-eliciting rhetoric about immigration may mirror the effects shown in research concerning print media, or, because it is an arena for more voices and where people with divergent opinions may or may not engage with each other, it could operate differently.

The current study aimed to fill the gap in the literature by analyzing immigration-relevant language in Twitter data. Specifically, we sought to answer the following two questions: (1) How prevalent was dehumanizing or disgust-eliciting language discussing immigrants or immigration on Twitter?; and (2) Did exposure to dehumanizing or disgust-eliciting language toward immigrants on Twitter cause other users to adopt negative attitudes on the issue? Due to the vast documentation of such language on traditional media (Marshall & Shapiro, 1972; O’Brien, 2003; Rozin et al., 2018) and preliminary evidence that it is also used online (e.g., Dhuman Giron & Vargas, 2020; Musolff, 2015), we expected this language to be occurring on Twitter. In addition, we expected that, consistent with prior literature showing effects of exposure to dehumanizing and disgust-eliciting language via traditional media (Dalsklev & Rønningsdalen-Kunst, 2015; Esses et al., 2013; Hodson et al., 2014; Marshall & Shapiro, 2018; Utych, 2018), exposure to language that dehumanizes or elicits a disgust response about immigrants or immigration on Twitter would negatively impact other users’ immigration attitudes. It is necessary to understand how this operates on social media because users may depict immigrants as subhuman, which could lead to society stigmatizing them and rejecting their uniquely human attributes (Musolff, 2015; Haslam, 2006). If people come to think of members of a group as objects, then they may think those group members have no “identity,” and, consequently, feel less empathy and compassion for those group members, conceptualize them as undeserving of the rights afforded to humans, and treat them without regard to societal standards (see Hodson et al., 2014). Social
media users who refer to immigrants with disgust-eliciting language could increase others’ negative attitudes towards immigrants, too, especially considering the evolutionary connection between disgust and dehumanization (Harris & Fiske, 2006). Testing these possibilities is critical given that potential effects range from support for stricter immigration policies (Dalsklev & Rønningsdal-Kunst, 2015; Esses et al., 2013) to genocide (Kelman, 1973; Stanton, 2016).

2. Methods

We employed Python to call on Twitter’s1 application programming interface (API) and then scrape, or gather, data from 43,500 U.S. Twitter users derived from Barberá et al. (2015).2 Because Republicans’ views toward immigration tend to be more negative on average than Democrats’ views (Brader et al., 2008; Utych, 2018), we sought to balance our sample on political affiliation and control for it in analyses. Therefore, we used only a random subset of still-active users whose party registration Barbera et al. had confirmed based on publicly available voter registration records. Once identifying this subset, we collected users’ IDs, their tweets, and the dates of their tweets. This information was used to identify a subsample of “communicators” who had tweeted about immigration during an observation period that extended from January 2018 to April 2020. Once the final sample of communicators was generated, we then scraped data from the communicators’ followers, heretofore referred to as “consumers,” as they could have been exposed to the communicators’ tweets. The tweets of the

---

1 Since this research was conducted, Twitter was acquired by Elon Musk and rebranded as “X.” We refer to the platform as it was known at the time of our study.

2 Barberá et al. (2015) collected data from 3.8 million active Twitter users in the U.S. and estimated their ideological preferences and social-network structures. The researchers examined whether users selectively exposed themselves to ideologically similar others and whether ideological preferences predicted users’ social media behavior. Their results showed that users’ networks and behavior varied across political and nonpolitical issues, but across topics conservative users were more likely than liberal users to select into echo chambers.
communicators and consumers comprise the data that was analyzed to test whether communicators’ use of dehumanizing or disgust-eliciting language toward immigration impacted consumers’ attitudes toward immigration. Each of these steps is described in greater detail next.

2.1 Sample Generation

2.1.1 Communicators

To generate the sample of communicators, data from the 43,500 Twitter users derived from Barberá et al. (2015) were analyzed to determine which users had tweeted about immigration during the observation period. The GloVe model from Gensim, which uses word vectors, was used to identify the 200 words closest in meaning and space across Twitter to the terms “immigrants” and “immigration.” The 200 words generated by GloVe were then used as search terms to identify 705,039 immigration-relevant tweets from the 43,500 Twitter users. Next, three research assistants reviewed a random sample of 100 tweets to determine if they were in fact immigration relevant. The first author and research assistants resolved through discussion whether they were or not. Due to the high number of irrelevant tweets produced through this process, the first and second authors identified the words from the vector that generated the irrelevant tweets, discarded those terms from the search set, and re-filtered the data. Following three iterations of this process, the first and second authors determined that immigration-relevant tweets could be reliably yielded by searching for the following words: immigration, immigrant, immigrants, migrant, migrants, deport, deported, deportation, deporting, customs, ICE, citizenship, undocumented, and illegals. The phrases illegal alien and illegal aliens were included because GloVe does not produce phrases, but these phrases are also used to discuss immigrants and immigration. To ensure that this modified list of search terms yielded a consistently immigration-relevant pool of tweets, three research assistants coded another
subsample of 100 tweets and verified they were all immigrant relevant. Therefore, filtering through all 43,500 users’ tweets using the condensed list revealed 40,826 immigrant-relevant tweets from 5,104 users, representing 12% of the full sample of users.

Among the subsample of 5,104 users who had tweeted about immigration during the observation period, the mode of immigration-relevant tweets was one, the average number of immigrant-relevant tweets was eight, and the median was two. To increase the likelihood that the communicators exposed consumers to their beliefs about immigration, we established which 100 users, balanced across political affiliation, discussed immigration in their tweets the most frequently. Therefore, the 50 registered Democrat communicators and 50 registered Republican communicators, as classified by Barberá et al. (2015), who had tweeted about immigration the most frequently were included in our final sample. In total, these 100 communicators provided a sample of \( N = 8,623 \) tweets, which was 21% of the sample of immigrant-relevant tweets. Within this group, users tweeted immigrant-relevant tweets between 53 to 233 times over the two-year observation period, with an average of 87.80 (\( SD = 36.51 \)) immigration-relevant tweets per user. Democrat and Republican communicators did not differ significantly in the number of immigration-relevant tweets they posted during the observation period, \( t(98) = -1.31, p = .19\); Democrats posted an average of 81.44 tweets and Republicans posted an average of 91.02 tweets.

2.1.2 Consumers

After finalizing the subsample of communicators, we then scraped their followers’ user IDs. The communicators’ followers are henceforth referred to as the consumers. We used the same list of search terms described previously to then scrape the consumers’ immigration-relevant tweets. This generated a subsample of 2,110 consumers with a total of \( N = 30,857 \)
immigration-relevant tweets over the two-year sampling period. Within this group, users tweeted immigration-relevant tweets between 1 to 376 times, with an average of 14.67 ($SD = 26.56$) immigration-relevant tweets per user.

### 2.2 Independent Variables: Dehumanizing and Disgust-Eliciting Language

The independent variables of dehumanization and disgust-eliciting language were determined by machine learning models through the Gradient Boosting for Classification model from sklearn (Pedregosa et al., 2011). We trained the Gradient Boosting for Classification machine learning model on a subsample of tweets. The subsample dataset that resembled the full dataset was created by having three research assistants independently code a subsample of 1,200 immigration-relevant tweets as either (a) dehumanizing or not and (b) disgust-eliciting or not. Agreement was reached on 171 tweets that were dehumanizing and 174 tweets that were not, yielding a training set of 345 tweets to use to create a model that would identify dehumanizing language. Using the same random subsample of 1,200 tweets, the same three research assistants reached agreement coding 144 tweets that were disgust-eliciting and 72 tweets that were not. These tweets were used to compose the training dataset of 216 observations for generating a model to identify disgust-eliciting language.

These are not particularly large numbers of observations to train models with; however, the Gradient Boosting for Classification machine learning model does not require a lot of training data to predict accurately (Pedregosa et al., 2011). Further, to ensure a more accurate estimation, one can randomly divide the testing sample into train/test subsets several times and average the error over these divisions. Thus, we used 20 random splits to evaluate the two models. We also modified the parameters of each model based on the data so as to provide the highest level of accuracy possible. This resulted in separate models that could predict
dehumanizing language in tweets with 80% accuracy and disgust-eliciting language in tweets with 86% accuracy. After establishing these levels of accuracy, we re-ran the models on their respective trained datasets to ensure they accurately predicted dehumanization language at least 80% of the time and disgust-eliciting language in tweets at least 86% of the time. Finally, we ran the models on the communicators’ tweets to identify the frequency with which they used dehumanizing language or disgust-eliciting language in their tweets about immigration.

2.3 Dependent Variable: Sentiment

Attitudes toward immigration were measured by conducting a sentiment analysis on consumers’ tweets. Sentiment analysis has been established to be a reliable tool for determining attitudes or sentiment on Twitter (e.g., Arias, Arratia & Xuriguera, 2014; Mohammed, 2016). We imported a natural language processing library called Textblob into Jupyter Notebook, which has been shown to be more accurate and reliable in identifying the sentiment of tweets in comparison to other libraries (Hasan, Moin, Karim, & Shamshirband, 2018). A sentiment analysis was performed on the consumers’ cleaned data, identifying the sentiment of the tweets as either negative (-1), neutral (0), or positive (1).

2.4 Analytic Approach

Vector Autoregressive modeling (VAR) and post-estimation tests of Granger-causality were used to test whether exposure to dehumanization or disgust-eliciting language Granger-caused a change in consumers’ sentiment on immigration, and, if so, in what direction the sentiment changed. Each equation in a VAR model can be interpreted as an Ordinary Least Squares (OLS) equation, with coefficients providing insight about whether effects are positive or negative. The benefit of testing for Granger-causality is that it determines whether one series forecasts another (Sobel & Osoba, 2009). Because Granger-causality is not true causality (Sobel
ATTITUDES TOWARD IMMIGRATION

& Osoba, 2009), the models cannot be interpreted as causal functions; rather, they must be interpreted as “Ganger-cause” functions. The assumption of Granger-causality is as follows: if X predicts Y better than Y’s past predicted Y, then X Granger-causes Y. In this case, we are testing whether communicators’ dehumanizing or disgust-eliciting language predicts the sentiment of consumers’ tweets better than the sentiment of consumers’ own past tweets does.

Prior to running analyses, data were transformed to make it suitable for time-series analysis. Specifically, variables were collapsed by a specified time period. Taking into consideration the desire for analytic power based on a required minimum of 50 observations to estimate the coefficients in time-series analyses (Tabachnick & Fidell, 2013), collapsing the data by years or months was inadequate and instead weeks were chosen as the time interval. Therefore, the data ranged from the first week in January 2018 to the second week in April 2020 and comprised averaged frequencies and ratings per week for dehumanizing tweets about immigration from the communicators, disgust-eliciting tweets about immigration from the communicators, and sentiment of the consumers’ tweets about immigration, as well as the percent of communicators registered as Democrats versus Republicans. Averaging the data into weekly variables yielded a dataset of 119 observations per variable.

3. Calculation

The VAR models, which are the basis for our Granger-causality tests, are specified below in equations (1) through (10). The predictors are abbreviated as follows: “S” represents sentiment of the consumers’ tweets, “DH” represents the dehumanizing language of communicators’ tweets, “DG” represents the disgust-eliciting language of communicators’ tweets, and “PA” represents the political affiliation of the communicators. “X” represents the regression coefficient
of the subscripted values whereas “(t-1)” signifies the identified number of lags, “C” is the
intercept, and “e” is the residual of the subscripted value at time t.

In (1), the predictors are the lagged values of “S” and “DH” on “S” at time t. In (2), the
predictors are the lagged values of “S” and “DH” on “DH” at time t. The post-estimation
Granger-causality tests use this information to determine whether communicators’ dehumanizing
language Granger-causes sentiment in consumers’ tweets by testing the null hypothesis that the
estimated coefficients on the lagged value of “DH” are jointly zero when regressing “S” on its
own lagged values and on the lagged values of “DH.”

\[ Y_{St} = C_S + X_{1,1} Y_{S(t-1)} + X_{1,2} Y_{DH(t-1)} + e_{St} \]  \hspace{1cm} (1)

\[ Y_{DH,t} = C_{DH} + X_{2,1} Y_{DH(t-1)} + X_{2,2} Y_{S(t-1)} + e_{DH,t} \]  \hspace{1cm} (2)

In (3), the predictors are the lagged values of “S,” “DH,” and “PA” on “S” at time t. In
(4), the predictors are the lagged values of “S,” “DH,” and “PA” on “DH” at time t. The post-estimation
Granger-causality tests use this information to determine whether communicators’ dehumanizing
language Granger-causes sentiment in consumers’ tweets while controlling for communicators’
political affiliation by testing the null hypothesis that the estimated coefficients on the lagged
value of “DH” and “PA” are jointly zero when regressing “S” on its own lagged values and on
the lagged values of “DH” and “PA.”

\[ Y_{S,t} = C_S + X_{1,1} Y_{S(t-1)} + X_{1,2} Y_{DH(t-1)} + X_{1,3} Y_{PA(t-1)} + e_{S,t} \]  \hspace{1cm} (3)

\[ Y_{DH,t} = C_{DH} + X_{2,1} Y_{DH(t-1)} + X_{2,2} Y_{S(t-1)} + X_{2,3} Y_{PA(t-1)} + e_{DH,t} \]  \hspace{1cm} (4)

\[ Y_{PA,t} = C_{PA} + X_{3,1} Y_{PA(t-1)} + X_{3,2} Y_{DH(t-1)} + X_{3,3} Y_{S(t-1)} + e_{PA,t} \]  \hspace{1cm} (5)

In (6), the predictors are the lagged values of “S” and “DG” on “S” at time t. In (7), the
predictors are the lagged values of “S” and “DG” on “DG” at time t. The post-estimation
Granger-causality tests use this information to determine whether communicators’ disgust-eliciting language Granger-causes sentiment in consumers’ tweets by testing the null hypothesis that the estimated coefficients on the lagged value of “DG” are jointly zero when regressing “S” on its own lagged values and on the lagged values of “DG.”

\[ Y_{S,t} = C_S + X_{1,1}Y_{S(t-1)} + X_{1,2}Y_{DG(t-1)} + e_{S,t} \]  
\[ Y_{DG,t} = C_{DG} + X_{2,1}Y_{DG(t-1)} + X_{2,2}Y_{S(t-1)} + e_{DG,t} \]

In (8), the predictors are the lagged values of “S,” “DG,” and “PA” on “S” at time \( t \). In (9), the predictors are the lagged values of “S,” “DG,” and “PA” on “DG” at time \( t \). Finally, in (10), the predictors are the lagged values of “S,” “DG,” and “PA” on “PA” at time \( t \). The post-estimation Granger-causality tests use this information to determine whether communicators’ disgust-eliciting language Granger-causes sentiment in consumers’ tweets while controlling for communicators’ political affiliation by testing the null hypothesis that the estimated coefficients on the lagged value of “DG” and “PA” are jointly zero when regressing “S” on its own lagged values and on the lagged values of “DG” and “PA.”

\[ Y_{S,t} = C_S + X_{1,1}Y_{S(t-1)} + X_{1,2}Y_{DG(t-1)} + X_{1,3}Y_{PA(t-1)} + e_{S,t} \]  
\[ Y_{DG,t} = C_{DG} + X_{2,1}Y_{DG(t-1)} + X_{2,2}Y_{S(t-1)} + X_{2,3}Y_{PA(t-1)} + e_{DG,t} \]  
\[ Y_{PA,t} = C_{PA} + X_{3,1}Y_{PA(t-1)} + X_{3,2}Y_{DG(t-1)} + X_{3,3}Y_{S(t-1)} + e_{PA,t} \]

4. Results

4.1 Frequency of Dehumanizing and Disgust-eliciting Language in Communicators’ Immigration-Relevant Tweets

We first examined the extent to which communicators use dehumanizing and disgust-eliciting language when tweeting about immigration, and found it was prevalent. Specifically, 45% (\( n = 3,197 \)) of communicators’ immigration-relevant tweets included dehumanizing
language and 50% ($n = 4,329$) included disgust-eliciting language. A chi-square goodness-of-fit test revealed that significantly more tweets about immigration did not include dehumanizing language than did, $\chi^2(1, N = 8,623) = 72.01, p < .05, \phi = .09$. However, a second chi-square goodness-of-fit test determined that tweets did not significantly differ in their likelihood of including disgust-eliciting language in relation to immigration, $\chi^2(1, N = 8623) = .159, p < .05, \phi = .004$, indicating that immigration-relevant tweets were just as likely as not to be disgust-eliciting.

4.2 Does Communicators’ Use of Dehumanizing or Disgust-eliciting Language toward Immigration Granger-Cause Consumers’ Sentiment?

Next, we investigated whether one’s use of dehumanizing or disgust-eliciting language toward immigrants on Twitter Granger-causes others to adopt negative attitudes on the issue. As shown in Table 1, the coefficients from the VAR equation suggest that, contrary to our expectations, communicators’ use of dehumanizing language in tweets about immigration was significantly and positively related to the sentiment of consumers’ immigration-relevant tweets. Specifically, every one unit increase in communicators’ average use of dehumanizing language per week was associated with a positive increase in consumers’ average sentiment about immigration per week by $.059$ units. That is, as communicators used more dehumanizing language in their immigration-relevant tweets, consumers’ attitudes on the issue became more favorable. Moreover, the $R^2$ from the VAR model suggests that communicators’ dehumanizing language about immigration accounted for 13% of the variation in consumers’ sentiment. Even so, the Granger-causality test showed that the communicators’ average use of dehumanizing language in tweets about immigration did not Granger-cause the sentiment expressed in consumers’ tweets, as communicators’ dehumanizing language did not predict consumers’ later
sentiment about immigration in tweets better than their baseline level of sentiment did, $F(3, 116) = 2.170, p = .096$. Thus, despite the variables being positively associated with one another, these results suggest the relation was not causal in nature.

With regard to use of disgust-eliciting language, as shown in Table 2, the coefficients from the VAR equation suggest that the communicators’ use of this type of language when tweeting about immigration is significantly and positively related to the sentiment expressed in consumers’ immigration-relevant tweets. Every one unit increase in communicators’ average use of disgust-eliciting language per week was associated with a positive increase in the average sentiment of consumers’ immigration-relevant tweets per week by .053 units. Again, this effect was not in the expected direction. The $R^2$ from the VAR model suggests that communicators’ disgust-eliciting language accounted for 8% of the variation in consumers’ sentiment toward immigration. However, there was no evidence of Granger-causality for the model analyzing the impact of communicators’ disgust-eliciting language on consumers’ sentiment, $F(3, 116) = 1.965, p = .124$. Therefore, again, communicators’ disgust-eliciting language did not predict consumers’ later sentiment about immigration in tweets better than their baseline level of sentiment did, and, despite the variables being associated with each other, the relation does not appear to be causal in nature.

4.3 Replicating Analyses While Accounting for Political Affiliation

Past research has shown relations between political affiliation and attitudes toward immigration, such that Republicans tend to have less favorable views as compared to Democrats (Brader et al., 2008; Utych, 2018). Therefore, we examined whether communicators’ political affiliation, as determined by Barberá et al. (2015), related to the frequency with which they used dehumanizing or disgust-eliciting language in their immigration-relevant tweets. As depicted in
Figure 1, we found that communicators who were registered Republicans tweeted a higher percentage of tweets that included dehumanizing language about immigration as compared to communicators who were registered Democrats. A chi-square test of independence showed that this difference was statistically significant, \( \chi^2(1, N = 8,623) = 253.78, p < 0.001, \phi = .17 \). In fact, 76\% (\( n = 2,435 \)) of the 3,197 immigration-relevant tweets that were labeled as dehumanizing came from communicators who were registered Republicans. In contrast, we found that registered Democrats tweeted a higher percentage of tweets that used disgust-eliciting language to discuss immigration as compared to registered Republicans (see Figure 1). Another chi-square test of independence determined the association between disgust-eliciting language in immigration-relevant tweets and political affiliation also was significant, \( \chi^2(1, N = 8,623) = 7.33, p = 0.007, \phi = .03 \).

Next, we examined the effects of communicators’ dehumanizing and disgust-eliciting language towards immigration on the sentiment of consumers’ tweets on the issue when controlling for communicators’ political affiliation.\(^3\) The pattern of findings differed in some respects from those reported previously.

To begin, as shown in Table 3, in the model for dehumanizing language, communicators’ political affiliation was significantly and positively associated with consumers’ sentiment in tweets related to immigration. Specifically, a one unit increase in the average number of communicators who were registered Democrats per week was associated with an increase in the

---

\(^3\) To ensure political affiliation could be included as a control variable in analyses, we checked the assumption of homoscedasticity by verifying that communicators’ political affiliation did not interact with the level of dehumanizing or disgust-eliciting language they exhibited in their immigration-relevant tweets to influence the sentiment of consumers’ tweets. Linear regression analyses indicated that neither of the interaction effects reached significance, \( \beta = 0.00, t(116) = 0.02, p = .99 \), and \( \beta = 0.26, t(116) = 1.40, p = .16 \), respectively.
average sentiment of consumers’ immigration-relevant tweets per week by .051 units. The $R^2$ from the VAR model suggests that communicators’ political affiliation explained 21% of the variation in consumers’ sentiment toward immigration. Further, the model showed that the lagged values of political affiliation of the communicators Granger-caused consumers’ sentiment in their later tweets about immigration even after controlling for percentage of Democrats among communicators at baseline, $F(3, 116) = 2.87, p = .040$. The direction of this effect is further validated by the fact that the sentiment expressed in consumers’ immigration-relevant tweets did not Granger-cause the communicators’ political affiliation, $F(3, 116)= .700 p = .554$.

In regard to the hypothesized effects, in the VAR model accounting for communicators’ political affiliation, communicators’ usage of dehumanizing language in immigration-relevant tweets was again significantly but unexpectedly positively associated with consumers’ sentiment related to immigration, just as when political affiliation was not included in the VAR model. Specifically, a one unit increase in communicators’ average use of dehumanizing language per week was associated with an increase in the average sentiment of consumers’ immigration-relevant tweets per week by .084 units. The $R^2$ from the VAR model suggests that communicators’ use of dehumanizing language explained 14% of the variation in consumers’ sentiment toward immigration. However, in contrast to the prior model, communicators’ use of dehumanizing language in tweets about immigration Granger-caused the sentiment expressed in consumers’ tweets when controlling for political affiliation, $F(3, 116)= 4.23, p = .007$. The direction of this effect is further validated by the fact that the sentiment expressed in consumers’ immigration-relevant tweets did not Granger-cause the communicators’ use of dehumanizing language in tweets, $F(3, 116)= .429, p = .732$. Thus, dehumanizing language influences
sentiment rather than sentiment influencing dehumanizing language, but the causal effect manifests only when accounting for communicators’ political affiliation.

With regard to the model involving disgust-eliciting language, as shown in Table 4, unlike in the dehumanization model, political affiliation was not significantly associated with consumers’ sentiment about immigration in tweets. Moreover, there is no evidence of Granger-causality, as communicators’ political affiliation did not predict consumers’ later sentiment about immigration in tweets better than the consumers’ baseline level of sentiment did, $F(3, 116) = .843, p = .474$.

Finally, after accounting for communicators’ political affiliation in the model, the association between the communicators’ use of disgust-eliciting language and the sentiment expressed in consumers’ tweets became nonsignificant. This suggests that the positive effect observed in the prior model of communicators’ disgust-eliciting language on the sentiment expressed in consumers’ tweets was driven in some way or overpowered by communicators’ political affiliation. There is also no evidence of Granger-causality for the model analyzing the impact of communicators’ disgust-eliciting language on consumers’ sentiment in their immigration-relevant tweets, $F(3, 116) = 1.97, p = .130$. Thus, the addition of political affiliation as a control variable in the model did not change the results of the Granger-causality analysis as it did in the model examining effects of dehumanizing language, and instead seems to have diminished the association altogether.

5. Discussion

Prior research has shown that being exposed to dehumanizing and disgust-eliciting language via traditional media such as newspapers can impact one’s attitudes toward immigrants and immigration (Buckels & Trapnell, 2013; Hodson & Costello, 2007; Utych, 2018). To our
knowledge, our study using data from Twitter is the first to examine these effects in the novel context of social media. As we describe next, our results were sometimes consistent with the extant literature but other times in contradiction to it.

To begin, we found that people who communicate on Twitter frequently use dehumanizing and disgust-eliciting language when tweeting about immigration. Twelve percent of users in our initial sample discussed immigration in their online communications during the two-year observation period. Because our sample was derived from Barberá et al.’s (2015) study that selected users on the basis of their tweets about a range of political and nonpolitical events, none of which included immigration, it suggests that many users discuss immigration on social media. Our analyses also showed that 45% of immigration-relevant tweets included dehumanizing language. Although a chi-square goodness-of-fit test revealed that tweets were significantly more likely to not include dehumanizing language than to include it, the problematic language still appears in a substantial portion of tweets. Moreover, 50% of immigration-relevant tweets in our sample included disgust-eliciting language; tweets were statistically just as likely as not to contain language that would elicit feelings of disgust toward immigration or immigrants.

Although our findings indicate that Twitter users often discuss immigration and immigrants with dehumanizing and disgust-eliciting terms, we did not find any evidence that this sort of language negatively impacted consumers’ attitudes about immigration. This is notable as the extant literature suggests that frequent use of dehumanizing and disgust-eliciting language is likely to have deleterious implications. As Schmitt (2005) discussed, consistent media references to people in terms of a specific ideology can lead to conceptual metaphors about the groups to which they belong, and continual depictions of immigrants that use dehumanizing and/or
disgust-eliciting language could cause those who consume the negative media to explicitly or implicitly conceptualize immigrants as dangerous and undeserving of the rights afforded to humans and, in turn, exclude immigrants from the “moral community” (Opotow, 1990). This type of moral exclusion is thought to be a contributor to hate crimes and other mass atrocities (Cao et al., 2023; Müller & Schwarz, 2023; Opotow, 1990). Our results did not identify any potential for such negative impacts, but it could be that people’s immigration attitudes are affected only immediately following exposure to dehumanizing and disgust-eliciting language. Therefore, as we assessed attitudes over two years, it could be that the negative impact dissipated over time, thus making it undetectable in our findings. Our results, though unanticipated, point to the possibility that the negative relations between media exposure to dehumanizing and disgust-eliciting language and immigration attitudes are not as strong as previously assumed, or, alternatively, that the variables may relate to each other differently on social media than traditional media. Social media is still a relatively new method of communication and we do not yet fully understand how its effects may differ from those of traditional media because users can interact with the information to which they are exposed. That is, social media users can interact actively with the producers of content, typically in real time (Rafaeli & Sudweeks, 1997), whereas arguably traditional media is consumed relatively passively. This could contribute to our finding different effects than expected based on previous studies of traditional media. Because our results contradict prior literature by showing that Twitter users respond to dehumanizing and disgust-eliciting language by developing more positive attitudes toward immigrants, future research is vital for understanding whether our findings are replicable and reflect a unique phenomenon attributable to the distinct nature of social media.
Our results also suggest that Twitter users do not exist solely in echo chambers with like-minded peers; rather, the findings support the notion that Twitter is a space for readily exchanging opposing viewpoints. Specifically, we found the communicators’ average use of dehumanizing language in tweets about immigration was positively associated with consumers’ later sentiment towards immigration, although support for Granger-causality emerged only when political affiliation was included in the model. Thus, our findings stand in contrast to the literature suggesting that Twitter users could be trapped in an echo chamber which may amplify and radicalize their previously existing opinions (de Saint Laurent et al., 2020). Rather, our results support the notion that social media, and Twitter in particular, is a platform for debate and a space where those with divergent opinions may correspond openly about controversial topics (Klein, 2019). It is possible that consumers may have been pushing back against communicators’ dehumanizing and disgust-eliciting Tweets by subsequently expressing more positive sentiment towards immigration. This could explain why we did not replicate prior research showing negative effects of such language on media consumers’ immigration attitudes (see Utych, 2018; Buckels & Trapnel, 2013; Hodson & Costello, 2007). That is, the voice afforded by online platforms may help to protect users from passively absorbing dehumanizing and disgust-eliciting language and incorporating negative ideas into their own attitudes about immigrants. Our results are also somewhat in line with those from Siapera et al.’s (2018) Twitter analysis, which showed that there were largely two frames that dominated the platform in regard to refugees. Specifically, Siapera et al. (2018) found that some tweets reflected a far-right perspective in which refugees were framed as terrorists or rapists, even using the hashtag “#rapefugees” to convey communicators’ feelings, but most tweets portrayed a humanitarian frame such that refugees were discussed in regard to human rights and aid. Because such divergent frames can
coexist on Twitter, consumers in our study may have responded to dehumanizing tweets about immigrants with positive messages because they were intentionally trying to counter communicators’ negative expressions, perhaps by highlighting positive features of immigration or by attempting to humanize immigrants. Although we did not code our sample of tweets to explore this possibility, future research should examine the exact content of social media users’ responses to negative portrayals of immigrants to determine whether and how they use the platform to object via counted messaging. Future work also should consider whether consumers’ expressions of positive sentiment about immigration have any impact on the prejudiced attitudes of the communicators who originally dehumanized immigrants.

Our results also suggest that exposure to dehumanizing language about immigration may impact users differently than exposure to disgust-eliciting language. Specifically, we found that communicators’ disgust-eliciting language had a significant positive effect on consumers’ sentiment in immigration-relevant tweets in the VAR model without political affiliation. Even so, further analysis showed that communicators’ use of disgust-eliciting language did not Granger-cause consumers’ sentiment about immigration, regardless of whether political affiliation was included in the model. This could suggest that, although there may be some overlap between the two constructs, they should remain distinct and operationalized separately as they may not relate to attitudes about immigration (or other issues) in identical ways. Alternatively, our findings could be due to the fact that there are many types of disgust as well as individual differences in sensitivity to disgust (Rozin et al., 1994; Haidt et al., 1997; Rozin et al., 2008), and this study did not account for either of these factors.

Other individual difference factors also may be important to consider in future efforts to understand the complex and interactive processes operating on social media. In particular, young
people comprise the largest faction of Twitter users (Auxier & Anderson, 2021). Yet, it is unclear how user age affects the way content is consumed. The average age of users on social media is especially notable because of its departure from the older demographic of traditional print media such as newspapers (Shearer & Matsa, 2018), on which prior studies of the relations between dehumanization, disgust, and immigration attitudes have focused. Due to this age difference, users’ tweets may reflect divergent values from older populations in regard to immigration (Dhuman Giron & Vargas, 2020). Thus, media may impact attitudes differently depending on both the type of media (social media, print, news, etc.) being consumed (Richardson, 2007) and the populations consuming it. Consistent with this, one study found that older Puerto Rican individuals typically viewed print sources of media and perceived a significantly different reception to their arrival in a Florida community as compared to younger Puerto Rican individuals who had received their news on social media (Dhuman et al., 2020). Specifically, the older Puerto Rican individuals perceived their arrival as beneficial to the community due to the printed media’s representation of them whereas the younger Puerto Rican individuals perceived a backlash upon their arrival because social media portrayed it as such. Thus, more research is needed to understand whether individual differences condition effects identified in prior studies using traditional media, and how different kinds of social media users are impacted by the information to which they are exposed.

One individual difference factor we did take into consideration was political affiliation. Specifically, we controlled for political affiliation because prior research has indicated that it relates to immigration attitudes (Chandler & Tsai, 2001). Our results suggest the association may be more complex than typically acknowledged. We found that communicators used negative language in their tweets about immigration to a different extent as a function of their political
affiliation, but whether either group used the language more or less depended on the type of language being considered. Specifically, compared to Republicans, Democrats issued fewer tweets that were classified as dehumanizing but more tweets that included disgust-eliciting language. Moreover, when we controlled for political affiliation in analyses, as noted previously, it impacted the effects of communicators’ dehumanizing and disgust-eliciting language on consumers’ subsequent sentiment in different ways. Indeed, the inclusion of political affiliation in the dehumanizing language model resulted in Granger-causality, whereas without it there was no evidence of causality.

The lack of causality in the model for dehumanizing language that did not include communicators’ political affiliation could have been due to the omitted variable bias. In other words, correlation does not prove causation, and one must account for all relevant variables to determine whether a causal effect exists in nature. Of course, there are many unknown characteristics of users on social media which makes controlling for potentially confounding variables difficult, but one that could be especially relevant in this context is consumers’ political affiliation. We were fortunate to have obtained our sample of communicators from Barberá et al. (2015), who successfully tackled the task of identifying and validating Twitter users’ political affiliation; however, we did not implement the same complicated method to assess the political affiliation of the consumers in our sample. Even so, it is plausible that doing so was unnecessary given that we controlled for the political affiliation of the communicators. That is, by accounting for communicators’ political affiliation we may have inadvertently also controlled for consumers’ political affiliation. This is because of political homophily, which is the tendency for people with similar political beliefs to form ties and follow each other on social media (Colleoni et al., 2014). Indeed, Colleoni et al. (2014) found that both Republican and Democrat Twitter
users exhibit a high rate of political homophily. Regardless, we did not expect consumers’ political affiliation to influence the relations between dehumanizing language and immigration attitudes or disgust-eliciting language and immigration attitudes. Even if consumers solely followed ideologically similar users due to confirmation bias (Del Vicario et al., 2017) or echo chambers (Lemieux & Regens, 2012)—which our findings suggest may not be the case as communicators’ negative dehumanizing tweets were followed by an increase in consumers’ positive tweets about immigration—past literature dictates that dehumanizing and disgust-eliciting metaphors impact people’s attitudes across the political spectrum (Hodson et al., 2014). Therefore, we would expect to find exposure to such negative language to have a negative impact on consumers’ sentiment toward immigration irrespective of consumers’ political affiliation. Still, this is an empirical question, and future research should account for both communicators’ and consumers’ political affiliation to better understand how the latter variable impacts the effects of social media exposure to certain types of language on immigration attitudes.

5.1 Limitations and Caveats

As with any research, there are a number of limitations that require the findings be interpreted with caution. First, the results must be considered in light of the way in which the variables were operationalized. The machine learning model for dehumanizing language was trained on a relatively small number of tweets, and there was no mechanism to highlight the specific word or phrase in each tweet that caused it to be categorized as dehumanizing. Rather, the model analyzed the entire text of the tweet to determine whether each tweet should be categorized as dehumanizing. It could be that the model learned not based on the exact text content that was considered dehumanizing but instead a negative attitude that was expressed
holistically in the tweet. Similar issues could have arisen in the machine learning model for
disgust-eliciting language, as the same process was used to generate the model. However, to
assess construct validity, two research assistants conducted a post-check on separate random
samples of 150 dehumanizing tweets and 150 disgust-eliciting tweets, as identified by their
respective models. The coders reached 81% and 82% agreement, respectively, indicating we can
have a high degree of confidence in the construct validity of our independent variables.

Also, with regard to sentiment analysis, it has been found to be highly accurate on
Twitter, particularly when run through the Textblob library (Hasan et al., 2018), but it could still
be limited in its ability to measure immigration attitudes. The algorithm operates by averaging
the sentiment of every word in a tweet to determine whether it is positive, neutral, or negative.
Therefore, a tweet that may appear to convey a negative attitude toward immigration may
mistakenly be coded as positive. For example, one of the consumers’ tweets that displayed an
anti-immigration attitude stated, “wish there was a way to deport them as far as I’m concerned,
they renounced their citizenship.” This tweet was coded as positive because it includes three
positive terms (“wish,” “concerned,” and “citizenship”) and only two negative terms (“deport”
and “renounce”). As this example demonstrates, differences between “pro-immigration” and
“anti-immigration” attitudes may not always align with “positive/negative/neutral” terms. To
assess the potential scope of this issue in our data, two research assistants conducted a post-check
of our dependent variable. Specifically, they reviewed a random set of 100 positive tweets and a
separate random set of 100 negative tweets to determine whether they agreed that the tweet was
displaying the sentiment with which it had been labeled. Across all tweets, the raters agreed that
the tweets were labeled accurately 83% of the time. They agreed 69% of the time that positive
tweets were accurately labeled and 97% of the time that negative tweets were labeled accurately.
Although this gives us confidence that our dependent variable was operationalized effectively, future researchers could consider measuring immigration attitudes in more direct and holistic ways, particularly as our data suggest their relations to dehumanizing language, disgust-eliciting language, and political affiliation may be more complex than previously assumed.

In addition, we did not determine whether the tweets in our sample referred to legal or illegal immigration. However, research suggests they are frequently conflated in the U.S. (Dick, 2011; Kinefuchi & Cruz, 2015). For example, when Flores and Azar (2023) asked White participants who comes to mind when they think about “typical immigrants currently living in the U.S.” (p. 2122), only 6% mentioned legality in their responses. However, data from a second sample showed that the majority of White participants believed most immigrants are in the U.S. illegally and the most common immigrant archetype was “a low-status, undocumented man from Latin America” (p. 2139). Moreover, this archetype predicted the most negative attitudes toward immigration relative to other archetypes, consistent with other work showing more negative attitudes toward immigrants who entered the U.S. by illegal rather than legal means (Esses, 2021). Considering those prior findings, it will be important for future work to explore how legality shapes the tone and effects of social media posts regarding this policy issue.

Finally, since this research was conducted, Twitter has transitioned into “X.” It is not yet known how this transition has changed users’ engagement with the platform, but early polling suggests that people may be using X less frequently and a quarter of users anticipate abandoning it altogether (Dinesh & Odabaş, 2023). It remains to be seen how users’ reactions to the new ownership and rebrand will affect their online discourse in general or about immigration specifically, or whether the findings reported herein will hold over time.
5.2 Conclusion

Exposure to negative phrases about immigration has been shown to negatively impact people’s attitudes toward immigration policies (Dalsklev & Rønningsdal-Kunst, 2015). More worrisome still, as Hodson et al. (2014, p. 267) explain, is that “representing others as less-than-human can have profound consequences like delegitimizing the target and removing them from protections otherwise afforded to ‘people.’” For centuries, mass violence has resulted from the consistent use of dehumanizing and disgust-eliciting metaphors about “others” (Galtung, 1990). Yet, our results contradict prior research suggesting that use of dehumanizing and disgust-eliciting metaphors to depict immigration in public discourse causes individuals to adopt negative connotations about immigration (e.g., Hodson et al. 2014). In fact, our findings suggest that exposure to dehumanizing (but not disgust-eliciting) language in the context of social media may have an unexpectedly positive impact on attitudes toward immigration that are expressed on social media. These contradicting sets of findings point to the crucial need for more research on this matter, especially in light of the prevalence with which dehumanizing and disgust-eliciting language appear in tweets about this policy issue.

At least one study has shown that humanizing outgroup members may help to reduce prejudice against them. Specifically, Costello and Hodson (2010) exposed participants to a mock journal article that explained that the human-animal divide is smaller than previously thought. Participants humanized immigrants significantly more after exposure to the article relative to before. This effect held even among highly prejudiced individuals. Further, displaying positive information about the benefits of immigration prominently could prevent extreme negative reactions to immigrants and refugees (Esses et al., 2013). To this point, Murrar and Brauer (2019) suggest that describing people through stories that humanize them may be the best way to
reduce prejudice and change attitudes without activating a person’s personal resistance to the stimuli. Citron and Norton (2011) recommended that social media platforms develop algorithms to ensure users who engage in hate speech come into contact with this positive countermessaging, but research also should investigate whether social media users who encounter negative posts about immigration are countermessaging intuitively and, if so, whether it is effective at reducing communicators’ dehumanization and disgust-based attitudes. Such work could represent a jumping off point for eradicating systemic biases against immigrants and other groups that plague our society.
Acknowledgements

We thank Dr. Pablo Barberá from University of Southern California for generously sharing his corpus of Twitter users for use in this research. We also are grateful for the suggestions and insight into how to improve this paper provided by Dr. Alissa Worden and Dr. Frankie Bailey from the University at Albany as well as Dr. Anamika Twyman-Ghoshal from Brunel University London. We also wish to recognize the significant contributions made to this project by Ashley Tatis, Camilla Marzella, Isabella De La Vega, Ashley Coakley, Kayla Ashchmutat, Victoria Giorgio, Healy Turesky, Carlie Cegielski, Kristale Abdulla, Manshuen Yu, Brianna McKernan, and Jayonna Treacy in the PULSE Lab at the University at Albany, all of whom assisted in operationalizing variables. We also acknowledge the funding provided for this project by the University at Albany’s Benevolent Association Research and Creative Activity Grant, without which the analyses for this paper would not have been possible.

Declaration of interest: The authors have no competing or incentivizing interests at stake in the results of this research. Questions may be directed to the corresponding author, Dr. Najdowski at cnajdowski@albany.edu.

References


Moore, S., & Mcferran, B. (2017). She said, she said: Differential interpersonal similarities predict unique linguistic mimicry in online word of mouth. *Journal of the Association for Consumer Research, 2*(2), 229-245. DOI:10.1086/690942


Table 1

*Associations Between Communicators’ Dehumanizing Language and Sentiment in Consumers’ Tweets in the Third Lag of the VAR Model*

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effects of Consumers’ Sentiment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...on Communicators’ Dehumanizing Language</td>
<td>.258</td>
<td>.382</td>
<td>.67</td>
<td>.501</td>
</tr>
<tr>
<td>...on Consumers’ Sentiment</td>
<td>.191</td>
<td>.091</td>
<td>2.10</td>
<td>.038</td>
</tr>
<tr>
<td><strong>Effects of Communicators’ Dehumanizing Language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...on Communicators’ Dehumanizing Language</td>
<td>.253</td>
<td>.113</td>
<td>2.24</td>
<td>.027</td>
</tr>
<tr>
<td>...on Consumers’ Sentiment</td>
<td>.059</td>
<td>.027</td>
<td>2.18</td>
<td>.031</td>
</tr>
</tbody>
</table>
Table 2

*Associations Between Communicators’ Disgust-eliciting Language and Sentiment in Consumers’ Tweets in the Third Lag of the VAR Model*

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects of Consumers’ Sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…on Communicators’ Disgust-eliciting Language</td>
<td>.200</td>
<td>.111</td>
<td>1.80</td>
<td>.074</td>
</tr>
<tr>
<td>…on Consumers’ Sentiment</td>
<td>.170</td>
<td>.091</td>
<td>1.88</td>
<td>.063</td>
</tr>
<tr>
<td>Effects of Communicators’ Disgust-eliciting Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…on Communicators’ Disgust-eliciting Language</td>
<td>.387</td>
<td>.392</td>
<td>.99</td>
<td>.325</td>
</tr>
<tr>
<td>…on Consumers’ Sentiment</td>
<td>.053</td>
<td>.026</td>
<td>2.06</td>
<td>.041</td>
</tr>
</tbody>
</table>
### Table 3

**Associations Between Communicators’ Political Affiliation and Dehumanizing Language and Sentiment in Consumers’ Tweets in the Third Lag of the VAR Model**

<table>
<thead>
<tr>
<th>Effect of Communicators’ Political Affiliation</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>…on Communicators’ Political Affiliation</td>
<td>.031</td>
<td>.108</td>
<td>0.29</td>
<td>.770</td>
</tr>
<tr>
<td>…on Communicators’ Dehumanizing Language</td>
<td>-.039</td>
<td>.088</td>
<td>.044</td>
<td>.660</td>
</tr>
<tr>
<td>…on Consumers’ Sentiment</td>
<td>.051</td>
<td>.020</td>
<td>2.55</td>
<td>.012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of Consumers’ Sentiment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>…on Communicators’ Political Affiliation</td>
<td>.638</td>
<td>.485</td>
<td>1.31</td>
<td>.192</td>
</tr>
<tr>
<td>…on Communicators’ Dehumanizing Language</td>
<td>.245</td>
<td>.122</td>
<td>2.00</td>
<td>.048</td>
</tr>
<tr>
<td>…on Consumers’ Sentiment</td>
<td>.141</td>
<td>.090</td>
<td>1.56</td>
<td>.122</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of Communicators’ Dehumanizing Language</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>…on Communicators’ Political Affiliation</td>
<td>.019</td>
<td>.151</td>
<td>0.13</td>
<td>.898</td>
</tr>
<tr>
<td>…on Communicators’ Dehumanizing Language</td>
<td>.293</td>
<td>.394</td>
<td>0.74</td>
<td>.458</td>
</tr>
<tr>
<td>…on Consumers’ Sentiment</td>
<td>.084</td>
<td>.028</td>
<td>3.00</td>
<td>.003</td>
</tr>
</tbody>
</table>
### Table 4

*Associations Between Communicators’ Political Affiliation and Disgust-Eliciting Language and Sentiment in Consumers’ Tweets in the Third Lag of the VAR Model*

<table>
<thead>
<tr>
<th>Effect of Communicators’ Political Affiliation</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>… on Communicators’ Political Affiliation</td>
<td>.004</td>
<td>.099</td>
<td>.040</td>
<td>.966</td>
</tr>
<tr>
<td>… on Communicators’ Disgust-eliciting Language</td>
<td>-.084</td>
<td>.084</td>
<td>-1.00</td>
<td>.322</td>
</tr>
<tr>
<td>… on Consumers’ Sentiment</td>
<td>.026</td>
<td>.019</td>
<td>1.35</td>
<td>.180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of Consumers’ Sentiment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>… on Communicators’ Political Affiliation</td>
<td>.622</td>
<td>.478</td>
<td>1.30</td>
<td>.196</td>
</tr>
<tr>
<td>… on Communicators’ Disgust-eliciting Language</td>
<td>.476</td>
<td>.406</td>
<td>1.17</td>
<td>.244</td>
</tr>
<tr>
<td>… on Consumers’ Sentiment</td>
<td>.137</td>
<td>.094</td>
<td>1.46</td>
<td>.147</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of Communicators’ Disgust-eliciting Language</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>… on Communicators’ Political Affiliation</td>
<td>.186</td>
<td>.137</td>
<td>1.36</td>
<td>.178</td>
</tr>
<tr>
<td>… on Communicators’ Disgust-eliciting Language</td>
<td>.238</td>
<td>.116</td>
<td>2.04</td>
<td>.044</td>
</tr>
<tr>
<td>… on Consumers’ Sentiment</td>
<td>.052</td>
<td>.027</td>
<td>1.95</td>
<td>.054</td>
</tr>
</tbody>
</table>
Figure 1

Percentage of Tweets Including Dehumanizing and Disgust-Eliciting Language as a Function of Communicators’ Political Affiliation

![Diagram showing the percentage of tweets including dehumanizing and disgust-eliciting language for Democrats and Republicans. Democrats have 36% dehumanizing tweets and 52% disgust-eliciting tweets, while Republicans have 54% dehumanizing and 48% disgust-eliciting tweets.](image-url)