The Political Effects of Language Corrections

Yamil Ricardo Velez*

Abstract

Despite growing conflicts over the words we use to describe political issues, we know surprisingly little about how "language policing" affects public opinion. In this paper, I explore whether language corrections (i.e., correcting people for using controversial terms and encouraging them to adopt new terms) shape intergroup attitudes. Using a novel experimental design, I find that language corrections increase the adoption of new political terms, are most successful when corrections are ideologically congruent, and have disparate impacts on intergroup attitudes – producing positive, negative, and null effects depending on the groups associated with the correction. Specifically, I find that correcting people for using "illegal" can move attitudes in a negative direction, despite increased adoption of the term "undocumented," whereas highlighting the omission of "radical Islamic terrorism" has the potential to shift attitudes in an anti-Muslim direction. These findings underscore the importance of language as a cause and consequence of our political divide.

*Yamil Ricardo Velez is an assistant professor of Political Science at George Washington University, 2121 I Street, Washington, DC 20052 (yrvelez@gwu.edu).

Political polarization continues unabated in the United States, producing substantial gaps in policy preferences between Democrats and Republicans at the mass and elite levels (McCarty, Poole, and Rosenthal 2016). However, polarization has not only influenced the strength and location of policy attitudes but also how voters and politicians discuss issues (Jensen et al. 2012). In contemporary politics, positions on abortion are dichotomized as "pro-life" and "pro-choice." The "death tax" has replaced the "estate tax" in the conservative lexicon. In cities, coalitions of renters and long-term residents have sounded the alarm against "gentrification" rather than "redevelopment." Debates over the use of controversial political terms are especially pronounced in contemporary discussions regarding immigration and terrorism. Liberal and conservative groups have sparred over the use of "illegal immigrant," with pro-immigrant organizations preferring "undocumented" or "unauthorized" as a more inclusive substitute to "illegal." In the case of terrorism, conservatives have promoted "radical Islamic terrorism" to emphasize linkages between Islam and acts of terror.

These practices on both sides of the aisle are predicated on the assumption that language is an essential tool for steering public opinion; an assumption that is supported by decades of research on framing effects (Iyengar and Simon 1993; Chong and Druckman 2007). Increasingly, however, political strategies concerning language are taking the form of specifying which words are admissible or inadmissible in political discussions and debates. In this paper, I refer to these strategies as language corrections. Language corrections encompass situations where individuals are reprimanded for using politically controversial words and encouraged to adopt new ways of discussing political issues or groups. Though popular discussions about language corrections center on the supposed rise of political correctness across college campuses (Wilson 1995; Lakoff 2000), political elites also engage in the practice, as evidenced by Republican critiques of President Obama for his avoidance of the term "radical Islamic terrorism" throughout his presidency (Healy and Haberman 2015).

Language corrections constitute an underexplored topic of political communication. Consistent with framing effects (Iyengar and Simon 1993), language corrections set the terms of the debate. However, they go a step further by demarcating which words are deemed appropriate and inappropriate in political discussions. Thus, in addition to altering the relevant

considerations voters recall when considering an issue, language corrections also have the potential to minimize the spread of specific considerations via interpersonal communication (Druckman and Nelson 2003). Moreover, in contrast to the ephemeral nature of issue frames, language corrections – if successful – could durably affect how individuals think about politics, as words become part and parcel with the considerations they conjure up in voters' minds (Pérez and Tavits 2016).

Despite the growing relevance of language corrections, we know very little about their effects on intergroup attitudes. I attempt to close that gap by drawing on persuasion research to set up expectations and constructing an experimental design that mimics some of the critical features of the phenomenon. Specifically, I develop an experiment that corrects participants for using a politically controversial term (e.g., illegal immigrant), suggests a substitute (e.g., undocumented immigrant), and encourages them to adopt the alternative term. Contrary to expectations derived from the persuasion literature, I find that language corrections significantly increase the adoption of new political terms. Moreover, I find that language corrections that are ideologically congruent are more likely to be successful. Concerning attitudes, however, language corrections have mixed effects. Highlighting the omission of "radical Islamic terrorism" increases anti-Muslim policy attitudes and affect. However, correcting people for the use of "illegal" does not affect pro-immigrant sentiment and can reduce support for pro-immigrant policies. These findings highlight the possibility that while these corrections might alter the words people use to describe politics, more research is needed to understand their attitudinal effects.

Language Gaps and Corrections

As the United States has grown more politically polarized, linguistic differences have become predictive of issue stances and group identity. Scholars have developed automated text analysis tools that accurately estimate whether a voter or politician is liberal or conservative based on linguistic patterns (Laver, Benoit, and Garry 2003; Sylwester and Purver 2015). These differences in speech can reflect different political priorities among elites (Grimmer 2016) and moral reasoning patterns among voters (Kraft 2018). Indeed, using open-ended response data

from the American National Election Study, Kraft (2018) shows that liberal citizens tend to discuss politics using words that emphasize harm reduction and fairness whereas conservatives use words that emphasize in-group loyalty. Taken together, the growing language gap across mass and elite levels suggests that liberals and conservatives are not only disagreeing over the substance of policies but forming unique political dialects.

The implications of the language gap are two-fold. First, substantial differences in how we discuss politics can increase miscommunication and animosity. When policies are described using terms that do not possess a shared meaning, this can introduce obstacles to understanding and reduce the possibility of compromise (Jensen et al. 2012). The larger the linguistic chasm between groups, the more we should expect partisans and ideologues to insist on using terms that are familiar and congenial. These kinds of interactions are unlikely to be amicable in our current polarized environment. Second, as language itself becomes politicized, words can become efficient markers of identity that signal in-group/out-group status and the extent to which an individual identifies with a particular group (Gumperz 1982).

These dual motivations – discussing political issues using standard terms and employing language as a way of reaffirming group identity – can give rise to language corrections. Language corrections proceed in the following manner: (1) A makes a statement using a politically controversial term, (2) B reprimands A for using the term, and (3) B encourages A to use a new term. Considering this abstracted situation, we can see that a language correction is invoked when a linguistic taboo is breached. Whether the motivation to correct is instrumental in terms of facilitating political conversations or a reflection of group identity, the ultimate objective of a language correction is to encourage the adoption of a new term. However, an additional consequence of a language correction could be attitude change, given that some people might accept or reject the correction and update group-relevant attitudes accordingly.

In recent years, language corrections have become increasingly prevalent in liberal and conservative circles. News organizations like the New York Times have received a substantial number of complaints from readers for their use of "illegal immigration," prompting them to rule out the use of "illegal" and "alien" as nouns in their articles (Hiltner 2017). Pro-immigration advocates have chided public officials for using the term, arguing that it "de-

humanizes and marginalizes the people it seeks to describe" (Vargas 2012). In addition, with the rise of social media, an internet "call-out culture" has developed – as users have employed online platforms such as Twitter to denounce racist and sexist language among elites and members of the public (Ahmad 2015). Although language corrections might appear to be a liberal phenomenon, conservatives have also used similar strategies in the context of terrorism. In the aftermath of the Pulse nightclub shooting in Orlando, then-candidate Donald Trump attacked President Barack Obama and presidential candidate Hillary Clinton for not saying radical Islamic terrorism, motivating them both to justify the omission in subsequent interviews.

Though disputes over language are increasingly common and visible, we know little about whether these strategies affect public opinion. Given that language corrections are often deployed as a way of promoting the use of certain words, it is important to assess whether these strategies result in the uptake of new terms to describe issues or groups. Moreover, we should also consider whether these corrections have secondary consequences for attitudes. In the following section, I draw upon research on persuasion and cognitive dissonance as a guide for understanding the effects of this increasingly ubiquitous phenomenon.

The Potential Effects of Language Corrections

Language Corrections and Behavior

According to the persuasion literature, language corrections should have limited success as persuasive messages. Studies have shown that when people perceive constraints on their behavioral freedoms, this can strengthen their attachment to those behaviors (Miron and Brehm 2006). According to O'Keefe (2004), this is due to the fact that restrictions on behaviors can increase anger toward the source of those constraints and promote the use of counterarguments. Psychological theories of reactance produce the expectation that people invested in the use of particular words will respond to language corrections by refusing to adopt the new term or doubling down on their use of the politically contentious word. Therefore, these theories indicate that language corrections could strengthen one's commitment to particular political terms.

Hypothesis 1 *On average, language corrections will not affect the adoption of new political terms.*

Though the average effects of language corrections on behavior should be minimal, they should also be conditional on whether the "corrector" and the person being corrected share a political identity. After all, attitude change is more likely when persuasive appeals come from ingroup rather than outgroup members (Mackie, Gastardo-Conaco, and Skelly 1992). Moreover, when groups are salient, arguments are seen as more persuasive, and participants demonstrate increased recall when those arguments come from ingroup versus outgroup members (McGarty et al. 1994). Based on these findings, language corrections should encourage the adoption of new terms when those terms are politically congenial. For example, persuasive appeals that involve using "undocumented" instead of "illegal" should be more effective among liberals than conservatives, given the greater tendency among liberals to express pro-immigrant attitudes.

Hypothesis 2 *Liberal (conservative) language corrections will have a stronger effect among liberals (conservatives) than conservatives (liberals).*

Language Corrections and Attitudes

Language corrections might also affect attitudes if individuals update their attitudes toward groups and policies associated with the correction. Given that language corrections will often involve an argument about why specific terms should be preferred over alternatives, frames are implicitly embedded in language corrections. Thus, it is possible that language corrections function like frames and shift attitudes by altering the considerations that voters recall when thinking about issues and groups (see Chong and Druckman (2007) for a review).

Hypothesis 3 *Language corrections will move group-relevant attitudes in the direction of the correction.*

However, in addition to their total effect on attitudes, the effects of language corrections might be especially pronounced among those who adopt these new political terms. As research on cognitive dissonance has found, individuals are often motivated to bring their atti-

tudes in line with their behavior (Bem 1967; Acharya et al. 2015). For instance, Acharya et al. (2015) highlight the possibility that violence itself might shape the development of negative attitudes as a way of minimizing cognitive dissonance. From this perspective, the adoption of a language correction should serve as a signal of one's attitudes, potentially strengthening the impact of the language correction on group-relevant attitudes.

Hypothesis 4 *Conditional on adopting a new political term, the effects of the language correction will be strengthened.*

Data and Methods

I examine these hypotheses using an online experiment where I manipulate exposure to language corrections, provide individuals with an opportunity to adopt a new political term, and measure subsequent intergroup attitudes. A total of 1,356 participants were recruited from the online crowdsourcing website Mechanical Turk.[1] Upon agreeing to participate in the study, participants were first asked to write a short essay about one of two topics: immigration or terrorism. These two issues were selected due to their prevalence in political discourse and as a way of exploring whether liberal and conservative language corrections differ in their effects. The order of the essays was randomized.

If participants mentioned the word "illegal" in their essay on immigration, they were randomly assigned to one of four conditions that varied whether they were exposed (not exposed) to a language correction and encouraged (not encouraged) to change their language.[2] Participants in the language correction condition received a message recommending the use of "undocumented" as a substitute for "illegal" due to its dehumanizing nature.[3] Those in the control condition received a neutral message encouraging them to proceed to the next section. On the next page, participants' essays were reproduced with a find and replace function

[1] Berinsky, Huber, and Lenz (2012) find that although the Mechanical Turk subject pool differs from a nationally representative sample in important ways, effect sizes obtained using Mechanical Turk samples are comparable to those obtained using national samples.

[2] This latter factor was manipulated to address the potential for demand effects.

[3] Though it would be possible to incorporate an additional condition that encourages people to use "illegal" instead of "undocumented," not one respondent in a pilot study (N = 500) used the term "undocumented" to describe immigrants.

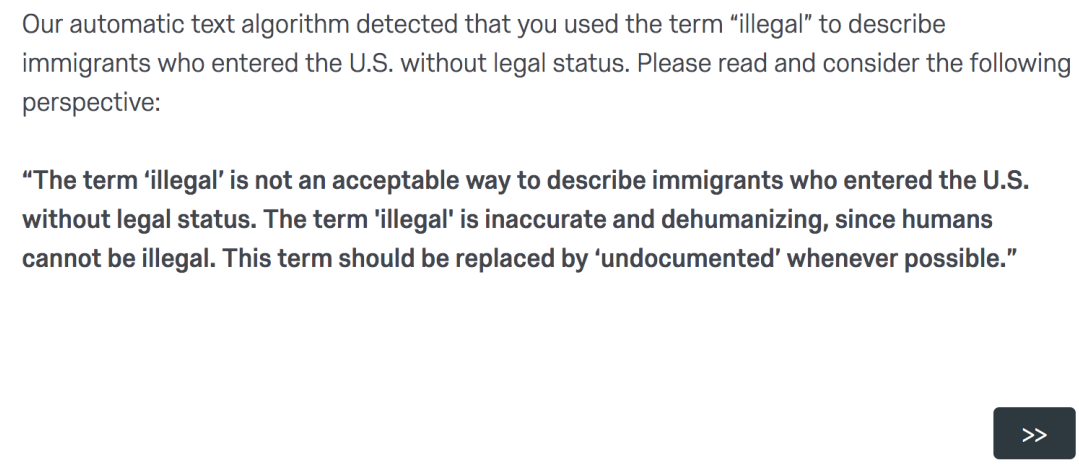

Figure 1: Reproduction of the language correction for those who mentioned “illegal” in their essay. See Appendix B for the terrorism language correction.

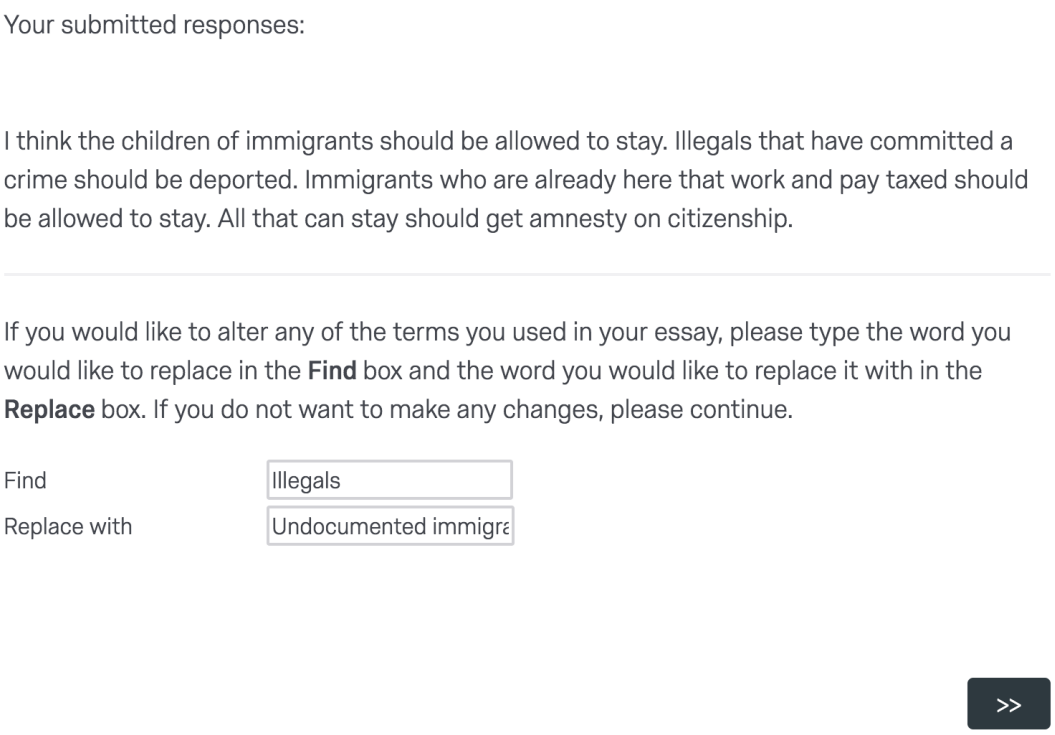

Figure 2: Example of the find and replace task for a subject in the no encouragement condition. See Appendix B for an example of the encouragement condition.

placed below the essay text. This task manipulated whether subjects were strongly or weakly encouraged to change their language. Participants were randomly assigned to either receive a statement encouraging them to change their essay using the find and replace function (i.e., by finding “illegal” and replacing it with “undocumented”), or a statement that implied that

they could change their text only if they wished to do so. Respondents were then taken to the questionnaire. If participants wrote about terrorism but failed to mention "radical Islamic terrorism," they were also randomly assigned to one of four conditions manipulating the language correction and encouragement. In this case, the language correction emphasized the omission of "radical Islamic terrorism" and distinguished it from other forms of terrorism. The encouragement condition was identical in both the immigration and terrorism cases. In the questionnaire portion of the study, all participants were asked about their attitudes toward immigrants and Muslims using measures of policy attitudes and affect. These particular items were included in a more extended set of questions that tapped into preferences and attitudes toward a variety of policies (e.g., economic policy) and groups (e.g., whites) in order to minimize experimenter demand. The study concluded with a measure of standard demographics and a debriefing.

Measures

I measured compliance with the language correction by using binary items that recorded whether subjects found the controversial term (i.e., "illegal" or "terror") and replaced it with the new term they were encouraged to adopt (i.e., "undocumented" or "radical"). Immigration policy attitudes were measured using a four-item summative scale that asked respondents about their preferred policies regarding undocumented immigrants (α = .84), whereas Muslim policy attitudes were measured using a three-item summative scale that captured the extent to which respondents would favor policies such as a ban on Muslim immigrants or protecting Muslims from hate crimes (α = .70).[4] These scales were coded such that higher scores indicated more pro-Muslim or pro-immigrant attitudes. Affect toward both groups was measured using a 100-point feeling thermometer scale.

Models

First, I estimate the average treatment effect of language corrections on compliance behaviors. That is, I examine whether issuing a language correction increased the likelihood of

[4]See supplemental appendix for question text.

a respondent changing the language in their essay. Second, I explore heterogeneous treatment effects by assessing whether liberals and conservatives respond differently to language corrections.[5] Third, I assess the attitudinal effects of language corrections by estimating the average treatment effect of language corrections on correction-relevant attitudes. Finally, I explore whether the adoption of a language correction strengthens the impact of the language correction on correction-relevant attitudes. In this case, I estimate a two-stage least squares regression where I use the language correction treatment as an instrument for whether the respondent adopted the correction. I control for age, ideology, partisanship, education, income, gender, and ethnic self-identification (i.e., Hispanic and Asian). In every analysis, two-tailed tests with $\alpha = .05$ are employed.

Results

Figure 3 displays the average treatment effect of language corrections on the likelihood of correcting one's essay. As shown in the first panel, a significant percentage of respondents complied with the language correction. Among those who omitted "radical Islamic terrorism," the language correction increased their propensity to include the term in their essay by about 10 percentage points ($\pm$ 3 percentage points). Among those who used the term "illegal," the language correction increased the likelihood of using "undocumented" by about 36 percentage points ($\pm$ 8 percentage points). Even though persuasion would seem difficult in this case, subjects did alter their language in response to the correction, and thus, I fail to find support for Hypothesis 1.

As shown in the second panel, the language correction increased the likelihood of using "radical Islamic terrorism" by about 16 percentage points for conservatives and 6 percentage points among liberals. Concerning immigration, the language correction increased the use of "undocumented" by about 49 percentage points among liberals and 21 percentage points among conservatives. These differences are statistically significant in both contexts ($p < .05$). Thus, there is evidence that politically congenial language corrections were more effective,

[5] I define liberals (conservatives) as those who score < 4 (> 4) on a 7-point ideology scale that runs from liberal and conservative.

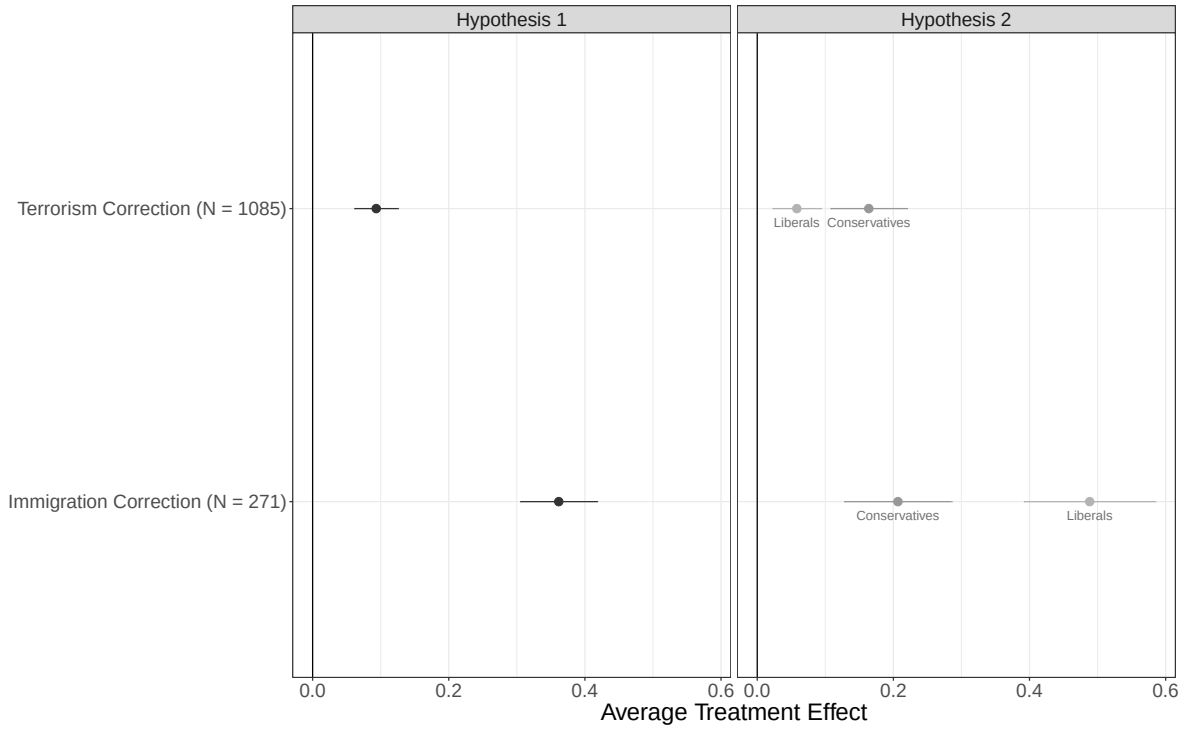

Figure 3: This figure presents point estimates and corresponding 95% confidence intervals for the average treatment effect (ATE) of language corrections on compliance with the language correction.

which is consistent with Hypothesis 2.

The first panel of Figure 4 displays the total effect of language corrections on policy attitudes and group affect. Among those who fail to mention "radical Islamic terrorism," receiving a language correction decreases pro-Muslim policy attitudes and affect by about 15 percent of a standard deviation ($\pm$ 12 percent of a SD). However, among those who describe immigrants as "illegal," receiving a language correction does not affect pro-immigrant affect and reduces support for pro-immigration policies by about 21 percent of a standard deviation ($\pm$ 20 percent of a SD). Therefore, in the case of terrorism, language corrections seem to move attitudes in the correction-intended direction whereas, in the case of immigration, corrections have no effect on pro-immigrant sentiment with the possibility of a backfire effect concerning policy attitudes. Thus, support for Hypothesis 3 is mixed.[6]

The second panel of Figure 4 displays complier average causal effects (CACE) for those who adopted the correction. Among those who adopted the correction in the terrorism context,

[6]I also explored whether these results were conditional on ideology. However, as shown in Appendix D, I fail to find support for conditional effects.

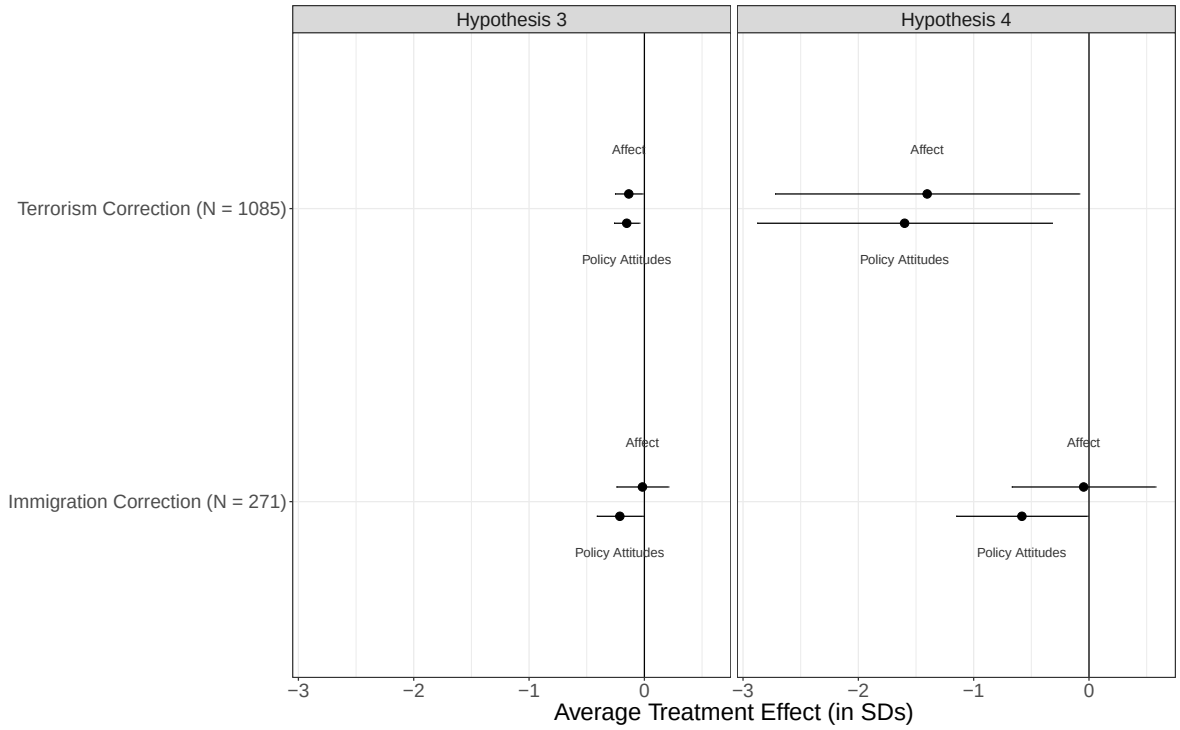

Figure 4: This figure displays the average treatment effects of language corrections (Facet 1) and complier average causal effects (Facet 2) with corresponding 95% confidence intervals.

the language correction moves attitudes in an anti-Muslim direction by about 1.5 standard deviations ($\pm$ 1.24 standard deviations). In the context of immigration, the language correction does not affect pro-immigrant sentiment but decreases support for pro-immigrant policies by about 58 percent of a standard deviation ($\pm$ 56 percent of a SD). Thus, Hypothesis 4 – the notion that language corrections are strengthened among those who adopt a new political term – obtains some support. Still, it is important to highlight that an adverse effect on immigration policy attitudes was observed, even among those who received the "undocumented" correction and agreed to update their language as a result.

Demand Effects

Although I found support for the notion that language corrections increase the adoption of new political terms and have mixed effects on attitudes, the key findings could be driven by demand effects. In a recent study of over 12,000 respondents using five distinct designs, Mummolo and Peterson (2018) fail to find support for demand effects using Mechanical Turk even when financial incentives are involved. Thus, despite common concerns about experimenter

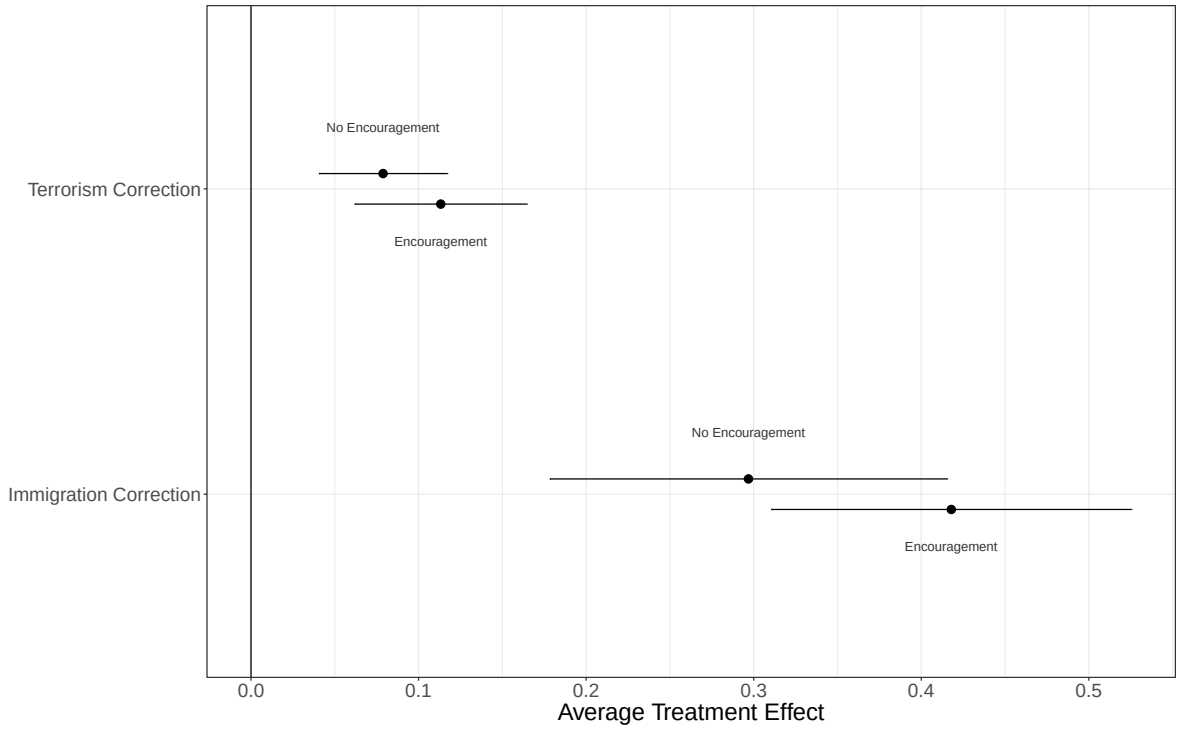

Figure 5: This figure displays the ATE of each correction, conditional on different encouragement conditions, with corresponding 95% confidence intervals.

demand on platforms such as Mechanical Turk, there is little evidence to suggest that they alter estimated treatment effects. Still, the experimental design included an additional condition that manipulated whether respondents were encouraged or not encouraged to adopt the new political term. As shown in Figure 5, though encouraging respondents to adopt the new political term increased compliance with the language correction, these differences are not statistically significant ($p > .05$). Given that this was the only task where respondents were explicitly asked to respond in a specific fashion, failure to find demand effects in this context strengthens the credibility of the findings.

Discussion

The findings presented here suggest that language corrections can exert powerful effects on behaviors and attitudes. In both immigration and terrorism contexts, language corrections increased the likelihood of adopting new political terms regardless of ideology. Still, there were ideological differences, such that liberals were more likely to adopt "undocumented" as a substitute for "illegal" whereas conservatives were more likely to use "radical Islamic

terrorism" when its omission was highlighted. In the case of terrorism, the introduction of "radical Islamic terrorism" had consistent negative effects on policy attitudes and affect. However, immigration attitudes were much less responsive to language corrections. In fact, while more people in the study adopted the use of "undocumented" than "radical Islamic terrorism," there was no corresponding positive movement in attitudes. It is possible that for words such as "radical Islamic terrorism" that are not a central part of the political lexicon, exposure to language corrections can move attitudes, much like an issue frame. However, when these words already guide partisan discourse on the topic, people might adopt new terms without changing their underlying attitudes toward target groups.

Despite growing contestation over language, we know surprisingly little about how the adoption of new words to describe social groups shapes intergroup attitudes and policy preferences. In this study, I found that language corrections can have dramatic effects on behavior, intergroup attitudes, and group-relevant policies. Future studies should explore features of terms that are most responsive to language corrections, and individual-level factors that contribute to the adoption of new political terms. Understanding the role of language in American politics could provide insights into whether growing conflicts over political language can exacerbate polarization and the conditions under which more inclusive language can improve intergroup relations.

Planned Extensions

Source variation

In the existing study, the experimenter is the source of the language correction. Aside from the potential issue of demand effects, previous research has highlighted the importance of source cues with respect to persuasion (Petty and Cacioppo 1984). Given that the ultimate goal of a language correction to persuade someone to adopt a new behavior, fixing the source limits what we can learn from the design. Therefore, a possible extension of the design would involve varying the identity of the individual issuing the correction. The expectation is that language corrections will exert the strongest effects when the source of the correction shares

an identity with the respondent (see Hypothesis 2).

Exploring less politicized language corrections

The most consistent effects of language corrections were observed in the terrorism condition. Though the purpose of this condition was to assess language corrections that appear on the right, controversies over the use of "radical Islamic terrorism" are largely an elite level phenomenon, and likely do not reflect how people talk about the issue. In contrast, the terms "illegal" and "undocumented" are routinely used to discuss the issue of immigration, and respondents might be more resistant to changing their attitudes in response to the language correction. An extension of the current design could correct respondents for terms that are not commonly used to describe political issues and evaluate whether language corrections are stronger under these conditions.

Language corrections based on omissions

An important difference between the immigration and terrorism conditions is that while respondents are corrected for using the word "illegal" in the immigration condition, those in the terrorism condition are corrected for omitting "radical Islamic terrorism." Respondents might be more amenable to changing their language in the latter condition because they might interpret this omission as an error. The existing immigration condition could be amended by including a correction that encouraging the respondent to also mention undocumented immigrants rather than asking them to substitute the word "illegal."

Improving ecological validity

Munger (2017) assesses the effects of reprimanding Twitter users for using racial slurs on the subsequent use of these terms, and finds that the prevalence of racial slurs over time declines when the source of the correction is a high status, in-group member. This design could be adapted to assess whether the relationships found in the Mechanical Turk study extend to a context where demand effects are less likely to be pertinent. Thus, an adaptation of the existing design would involve finding social media users who use the terms "illegal"

or “undocumented,” issuing a language correction, and assessing whether these corrections have subsequent effects on the use of these terms in the future.

References

Acharya, Avidit, Matthew Blackwell, Maya Sen et al. 2015. "Explaining attitudes from behavior: A cognitive dissonance approach." *Journal of Politics* .

Ahmad, Asam. 2015. "A note on call-out culture." *Briarpatch Magazine* 2.

Bem, Daryl J. 1967. "Self-perception: An alternative interpretation of cognitive dissonance phenomena." *Psychological review* 74(3): 183.

Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political Analysis* 20(3): 351–368.

Chong, Dennis, and James N Druckman. 2007. "Framing theory." *Annu. Rev. Polit. Sci.* 10: 103–126.

Druckman, James N, and Kjersten R Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47(4): 729–745.

Grimmer, Justin. 2016. "Measuring Representational Style in the House: The Tea Party, Obama, and Legislators' Changing Expressed Priorities.".

Gumperz, John J. 1982. *Language and social identity.* Vol. 2 Cambridge University Press.

Healy, Patrick, and Maggie Haberman. 2015. "95,000 words, many of them ominous, from Donald Trump's tongue." *The New York Times* 5.

Hiltner, Stephen. 2017. "Illegal, Undocumented, Unauthorized: The Terms of Immigration Reporting." *The New York Times* .

Iyengar, Shanto, and Adam Simon. 1993. "News coverage of the Gulf crisis and public opinion: A study of agenda-setting, priming, and framing." *Communication research* 20(3): 365–383.

Jensen, Jacob, Suresh Naidu, Ethan Kaplan, Laurence Wilse-Samson, David Gergen, Michael Zuckerman, and Arthur Spirling. 2012. "Political polarization and the dynamics of political

language: Evidence from 130 years of partisan speech [with comments and discussion]." *Brookings Papers on Economic Activity* , 1–81.

Kraft, Patrick W. 2018. "Measuring Morality in Political Attitude Expression." *The Journal of Politics* 80(3): 000–000.

Lakoff, Robin Tolmach. 2000. *The language war*. Univ of California Press.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(2): 311–331.

Mackie, Diane M, M Cecilia Gastardo-Conaco, and John J Skelly. 1992. "Knowledge of the advocated position and the processing of in-group and out-group persuasive messages." *Personality and Social Psychology Bulletin* 18(2): 145–151.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches*. MIT Press.

McGarty, Craig, S Alexander Haslam, Karen J Hutchinson, and John C Turner. 1994. "The effects of salient group memberships on persuasion." *Small Group Research* 25(2): 267–293.

Miron, Anca M, and Jack W Brehm. 2006. "Reactance theory-40 years later." *Zeitschrift für Sozialpsychologie* 37(1): 9–18.

Mummolo, Jonathan, and Erik Peterson. 2018. "Demand Effects in Survey Experiments: An Empirical Assessment.".

Munger, Kevin. 2017. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39(3): 629–649.

O'Keefe, Daniel James. 2004. "Trends and prospects in persuasion theory and research." In *Readings in persuasion, social influence, and compliance gaining*. Pearson/Allyn and Bacon.

Pérez, Efrén O, and Margit Tavits. 2016. "Language shapes public attitudes toward gender equality.".

Petty, Richard E, and John T Cacioppo. 1984. "Source factors and the elaboration likelihood model of persuasion." *ACR North American Advances* .

Sylwester, Karolina, and Matthew Purver. 2015. "Twitter language use reflects psychological differences between democrats and republicans." *PloS one* 10(9): e0137422.

Vargas, Jose Antonio. 2012. "Immigration debate: The problem with the word illegal." *Ideas Immigration Debate The Problem with the Word Illegal Comments* .

Wilson, John K. 1995. *The myth of political correctness: The conservative attack on higher education.* Duke University Press.

Online Supplemental Appendix

A. Essay Prompts

Immigration

> Please write a few sentences about your thoughts and feelings regarding immigrants in the United States. Specifically, think about debates that are occurring today concerning the legal status of immigrants who crossed the border.

Terrorism

> Please write a few sentences in response to the following prompt: Do you consider terrorism a significant issue in the world today? Why or why not?

B. Experimental Conditions

Language Correction (Terrorism)

> Our automatic text analyzer detected that you did not use the terms "radical Islamic terrorists" or "radical Islamic terrorism" in your response. Please read and consider the following perspective:
>
> "It's important to acknowledge the seriousness of radical Islamic terrorism, since this is a major strand of terrorism. One should always use the terms 'radical Islamic terrorist' or 'radical Islamic terrorism' when discussing issues regarding terrorism."

Language Correction (Immigration)

> Our automatic text algorithm detected that you used the term "illegal" to describe immigrants who entered the U.S. without legal status. Please read and consider the following perspective:
>
> "The term 'illegal' is not an acceptable way to describe immigrants who entered the U.S. without legal status. The term 'illegal' is inaccurate and dehumanizing, since humans cannot be illegal. This term should be replaced by 'undocumented' whenever possible."

Control (Terrorism and Immigration)

> Please continue.

Encouragement (Terrorism and Immigration)

Your submitted response:

We **highly recommend** that you make some changes to your essay. If you would like to alter any of the terms you used in your essay, please type the word you would like to replace in the Find box and the word you would like to replace it with in the Replace box. If you do not want to make any changes, please continue.

Find:

Replace with:

No Encouragement (Terrorism and Immigration)

Your submitted response:

If you would like to alter any of the terms you used in your essay, please type the word you would like to replace in the Find box and the word you would like to replace it with in the Replace box. If you do not want to make any changes, please continue.

Find:

Replace with:

C. Measures

All items were measured using a five-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree."

Muslim Policy Attitudes

- The United States government should institute a ban on Muslims entering the country. (R)
- United States government officials should be more vocal in condemning anti-Muslim hate crimes and violence.
- The United States government should take a more active role in protecting Muslims from discrimination.

Immigration Policy Attitudes

- Present levels of immigration should be decreased. (R)
- The federal government should increase spending for border security. (R)
- Unauthorized immigrants should be granted an opportunity to stay.
- We should do a better job of providing immigrants with economic and educational opportunities in this country.

R denotes reversed items.

D. Conditional Effects

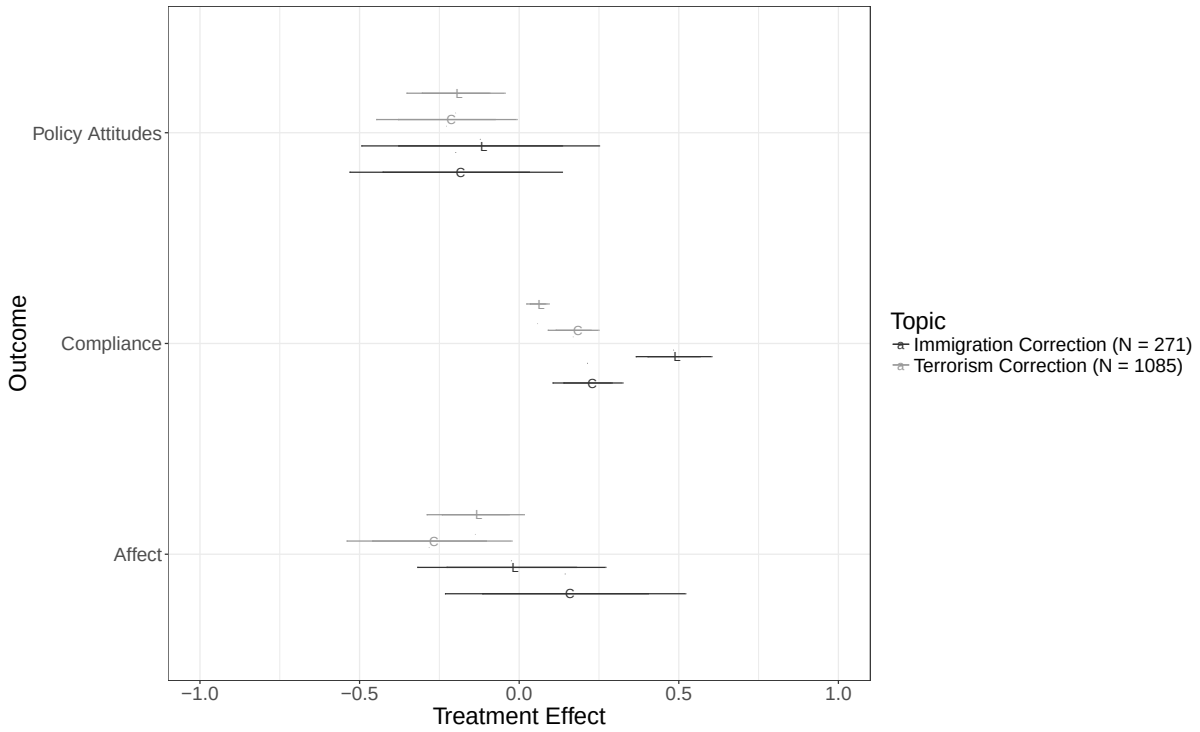

Figure 6: This figure displays the estimated effects of the language corrections among respondents who identify as liberals and conservatives. Though conservatives (liberals) are more likely to comply with the "radical Islamic terrorism" ("illegal") language correction, intergroup attitudes tend to move in the same direction. The most consistent findings are observed in the terrorism condition, where both liberals and conservatives express more anti-Muslim policy attitudes and affect.