Immigration policy agreement moderated through social media

Ian Vaimberg; Luis Cartagena; Alice Drozd; Timothy Majidzadeh; Edwin Fleurant

Introduction & Justification

As the 2024 U.S. presidential election unfolded, immigration remained a highly polarizing issue. Both major political parties sought voter approval by adopting increasingly hardline stances on border control and immigration policy. This raises a critical question: is a tough approach necessary to sway public opinion, or can more nuanced strategies—such as promoting pro-immigration sentiment—positively influence public approval of immigration?

Existing research suggests that public sentiment on contentious issues like immigration is not immutable. Political communication has long been recognized as a powerful driver of opinion formation, with political elites and media outlets shaping public attitudes through selective messaging (Zaller, 1992).[1] Empirical evidence from Canada, for instance, illustrates how pro-immigration messaging from government leaders during the 2010s shifted public perceptions, increasing warmth toward immigrants (Gaucher et al., 2018).[2] Similarly, Zhirkov, Verkuyten, and Ponarin (2021)[3] found that modest interventions emphasizing positive or negative portrayals of immigrant groups could significantly alter public sentiment. These findings suggest that public approval of immigration may reflect deeply held beliefs and respond to external messaging strategies.

Social influence, particularly in digital contexts, further supports this potential for opinion shaping. Bond et al. (2012)[4] demonstrated that spreading political mobilization via Facebook increased civic participation among direct recipients and their social networks. Fowler and Christakis (2010)[5] observed cooperative behaviors cascading through social networks, while Aral and Walker (2012)[6] showed how social validation cues, such as likes and comments, amplify susceptibility to influence. These studies highlight how social norms and peer validation can shape individual attitudes and behaviors, underscoring the decisive role of online networks in modern opinion formation.

Our study builds on these insights by investigating whether promoting pro-immigration sentiment using social validation cues can boost public approval of immigration. Specifically, we aim to test whether framing immigration-related statements as popular and socially endorsed—through indicators such as likes, shares, and supportive comments—will increase agreement with these statements among survey participants.

---

[1] *The Nature and Origins of Mass Opinion*, John R. Zaller, Cambridge University Press, New York, NY, 1992.
[2] Gaucher, D., Friesen, J., Neufeld, K., and Esses, V., “Changes in the Positivity of Migrant Stereotype Content: How System-Sanctioned Pro-Migrant Ideology Can Affect Public Opinions of Migrants,” *Social Psychological and Personality Science*, 9 (2) 2018, pp. 223-233.
[3] Zhirkov, K., Verkuyten, M., and Ponarin, E., “Social conformity and prejudice toward immigrants: the role of political messaging,” *Social Influence*, 16 (1) 2021, pp. 65-77.
[4] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H., “A 61-million-person experiment in social influence and political mobilization,” *Nature,* 489 (7415) 2012, pp. 295-298.
[5] Fowler, J. H., and Christakis, N. A., “Cooperative behavior cascades in human social networks,” *Proceedings of the National Academy of Sciences,* 107 (12) 2010, pp. 5334-5338.
[6] Aral, S., and Walker, D., “Identifying influential and susceptible members of social networks,” *Science*, 337 (6092) 2012, pp. 337-341.

We hypothesize that individuals exposed to pro-immigration messages accompanied by strong social validation will show higher levels of agreement than those exposed to the same messages without such cues. By employing a randomized controlled design with treatment and control groups, we aim to isolate the causal effects of social validation on opinion formation. Our findings will contribute to the growing body of research on social influence, offering new insights into the malleability of public opinion on immigration in a highly charged political context.

Methods

Comparison of Potential Outcomes

This study contrasts two potential outcomes: the level of agreement with pro-immigration messages among individuals exposed to social validation cues versus those who viewed the same messages without these cues. The comparison evaluates whether social validation enhances public approval of immigration messages, considering baseline agreement levels and variability across demographic and political subgroups.

A randomized controlled design was employed, assigning participants to either the treatment group (exposed to social cues) or the control group (no social cues). The analysis sought to isolate the causal impact of social validation, focusing on determining whether such cues significantly influence agreement with divisive pro-immigration statements. While the average treatment effect (ATE) was positive, statistical significance was not achieved ($p > 0.196$). This suggests that social validation cues may have a limited or inconsistent effect on opinion formation.

Furthermore, the results highlighted baseline differences in agreement across political affiliations but found no significant variation in treatment effects within these subgroups. This outcome underscores the complexities of opinion dynamics in politically polarized contexts, suggesting that external messaging strategies, such as elite-driven communication, might have a more pronounced influence than peer-level conformity within social media environments.

Randomization process

We partner with PureSpectrum,[7] a market research and insights platform, to conduct a stratified random sample of the U.S. adult population. We establish demographic quotas based on data from the 2020 U.S. Census[8] to ensure our sample of 200 completed survey responders matches the population in terms of political party, race, Hispanic or Latino origin, and the joint distribution of gender and age. Our respondents are directed to take a survey on the Qualtrics[9] platform. We screen each respondent for potentially fraudulent or duplicate responses and

---

[7] PureSpectrum Marketplace, available at www.purespectrum.com.
[8] "DP1 | Profile of General Population and Housing Characteristics," United States Census Bureau, 2020, available at https://data.census.gov/table?g=010XX00US&d=DEC+Demographic+Profile.
[9] Qualtrics, available at www.qualtrics.com.

collect their demographic profile information. We randomize respondents who meet our demographic quotas into equally sized treatment and control groups. We employ blocking and require that respondents are equally balanced between treatment and control based on political party, race, and Hispanic or Latino origin. This stratifying and blocking approach ensures that the sample is balanced between treatment and control for observed covariates.

Twenty-six respondents (twenty from the treatment group and six from the control group) did not complete the survey despite submitting their demographic information and seeing at least the first question. The individuals who attrited were replaced in the sample to ensure 200 completed responses. A Welch two-sample t-test shows that the attrition rate is statistically and significantly higher in the treatment group (p-value = 0.009). This is cause for concern; however, the attrition rate is not a statistically significant predictor of political party or demographic information. Additionally, among individuals completing at least one survey question, attrition status is not a statistically significant predictor of their response to the first question after controlling for treatment assignment and demographic covariates (p-value 0.123), and neither is the interaction term between treatment assignment and attrition (p-value 0.910). This reassures us that our randomization is sound despite the differential attrition rates.

Treatment

The treatment within this experiment was the administration of five divisive statements about immigrants and immigration policy with or without the presence of social authority. The five statements and their respective general categories were:

1. **Voting Rights:** “Despite immigrants' deep ties to their communities, it’s disgraceful that many lack the right to vote, relegating them to second-class citizens without proper representation.”
2. **Border Policy:** “Immigrants show courage and resilience by staking a path towards America. We should open our doors to all individuals coming here through traditionally legal or undocumented pathways.”
3. **Economic Benefits:** “It’s unjust that immigrants, who contribute so much to our country through their hard work and taxes, are often restricted and excluded from essential economic benefits like food stamps and unemployment insurance. They deserve access to the same support systems that help build a fair and equitable society.”
4. **Deportation Policy:** “Deporting immigrants is both morally wrong and a waste of limited government resources. Instead, we should implement mass amnesty and create an expedited pathway to citizenship, allowing immigrants to fully contribute to and call the U.S. their home.”
5. **Workers Rights:** “Undocumented workers are often exploited in the U.S., despite being a crucial part of the nation’s economy. To protect their rights and drive economic growth, we should take immediate steps to expand the availability of work visas significantly.”

These statements were curated to be both positive toward immigrants and immigration policy and divisive due to their somewhat extreme nature. Initially, our survey consisted of 10 less-divisive statements, but we found from the first pilot that the treatment effect was minimal. Reducing the number of survey statements did not reduce the treatment effect while also reducing the experiment's costs. As a result, we focused on making the statements more divisive and only selected five statements for the final survey.

The control group received these statements as is, with a white background, no additional visual effects, and in plain text. The treatment group received these statements as if they were posted to a social media platform by “The Civil Affairs Network,” a fictional political news entity (**Figure 1**). The “post” also included random reactions (e.g., likes, loves, surprised faces), comments, and shares. Additionally, to make the “post” believable, it included a few comments that agreed with the text in the post, also with a random number of reactions (likes or loves). We chose “Civil Affairs Network” because of its partisan ambiguity, meaning it did not inherently imply political lean either way.

Figure 1: Example of Experimental Treatment

The experiment was randomized on a person level tied to their IP address, so the same person could not receive the treatment and control. As such, we were not worried about any spillover effects.

Consort document

We collected 422 responses from the Qualtrics survey (**Figure 2**). Eight responses were excluded due to low ReCaptcha scores, high fraud scores, or duplicate responses. An additional 188 were excluded because we exceeded the quota of needed responses. After excluding those combined 196 responses, we were left with 226 randomized treatment and control responses. Of those 226, 120 were in the treatment group, and 106 were in the control group. Twenty treatment respondents did not complete their survey and were hence replaced. Similarly, six respondents from the control group did not complete their survey and were replaced. We included 100 treatment responses and 100 control responses in the experimental analysis.

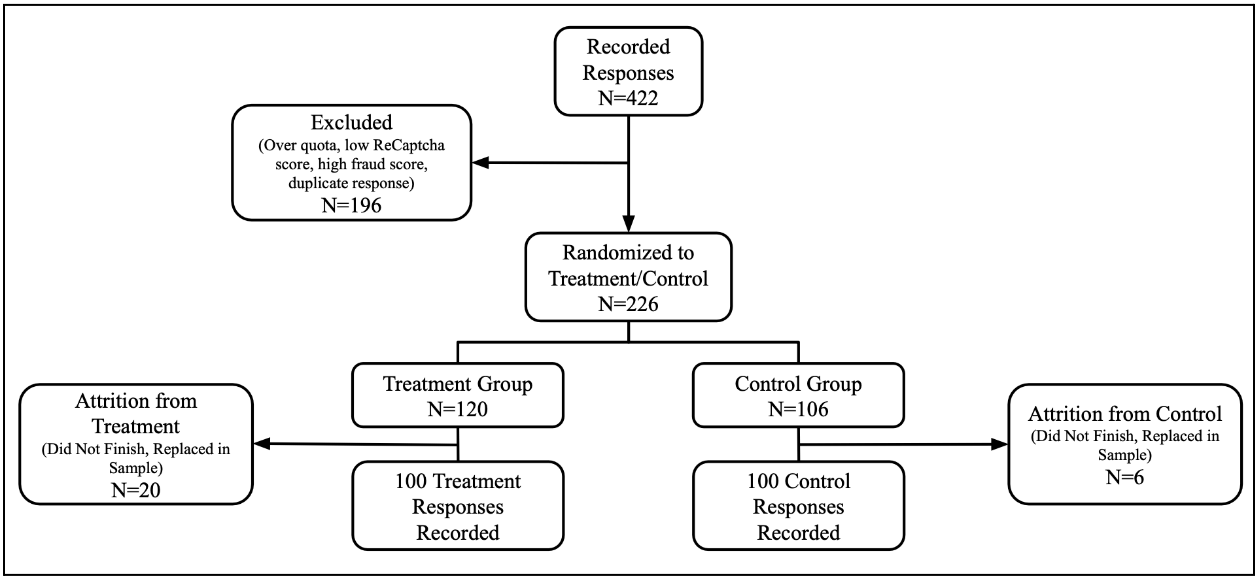

Figure 2: Cohort Flowchart

Power calculation: from pre-experiment analysis

Before the experiment was launched, we conducted a power analysis to estimate the necessary sample size to detect a meaningful treatment effect. This analysis, based on prior work, considered four scenarios:

1. **No Blocking, Uniform 4.3% Effect Size:** This baseline scenario did not include blocking and used a uniform treatment effect of 4.3% per question (0.434 on a 10-point scale), as informed by Williamson et al. (2021).[10]
2. **Blocking on Party, Uniform 4.3% Effect Size:** This scenario incorporated blocking by party affiliation (Democrat, Independent, Republican), recognizing significant differences in baseline attitudes toward immigration between these groups. This scenario also used a uniform treatment effect of 4.3%.
3. **Blocking on Party, Uniform 1.77% Effect Size:** This scenario, similar to scenario 2, also blocks by party affiliation, but it used a more conservative treatment effect of 1.77% (0.177 on a 10-point scale), reflecting an alternate survey design from Williamson et al. that showed a smaller effect.
4. **Blocking on Party, Non-Uniform Effect Size:** Here, treatment effects are not uniform across the political spectrum with the assumption of a +0.434 effect for Democrats, +0.177 for independents, and 0.000 for Republicans.

[10] Williamson, S., Adida, C., Lo, A., Platas, M., Prather, L., & Werfel, S. (2021). Family Matters: How Immigrant Histories Can Promote Inclusion. American Political Science Review, 115(2), 686-693.

Power estimates were computed using ordinary least squares (OLS) regression models. For scenario 1, the outcome variable of score is regressed on the treatment variable. For all other scenarios, score was regressed on the treatment and party affiliation of the participant. To determine the appropriate sample size to collect, sample sizes were tested from 25 to 500 individuals in increments of 25, with 100 simulations for each sample size. The resulting power of the simulated experiment was then recorded.

The resulting plot of power based on the simulation analysis is shown in the following graph:

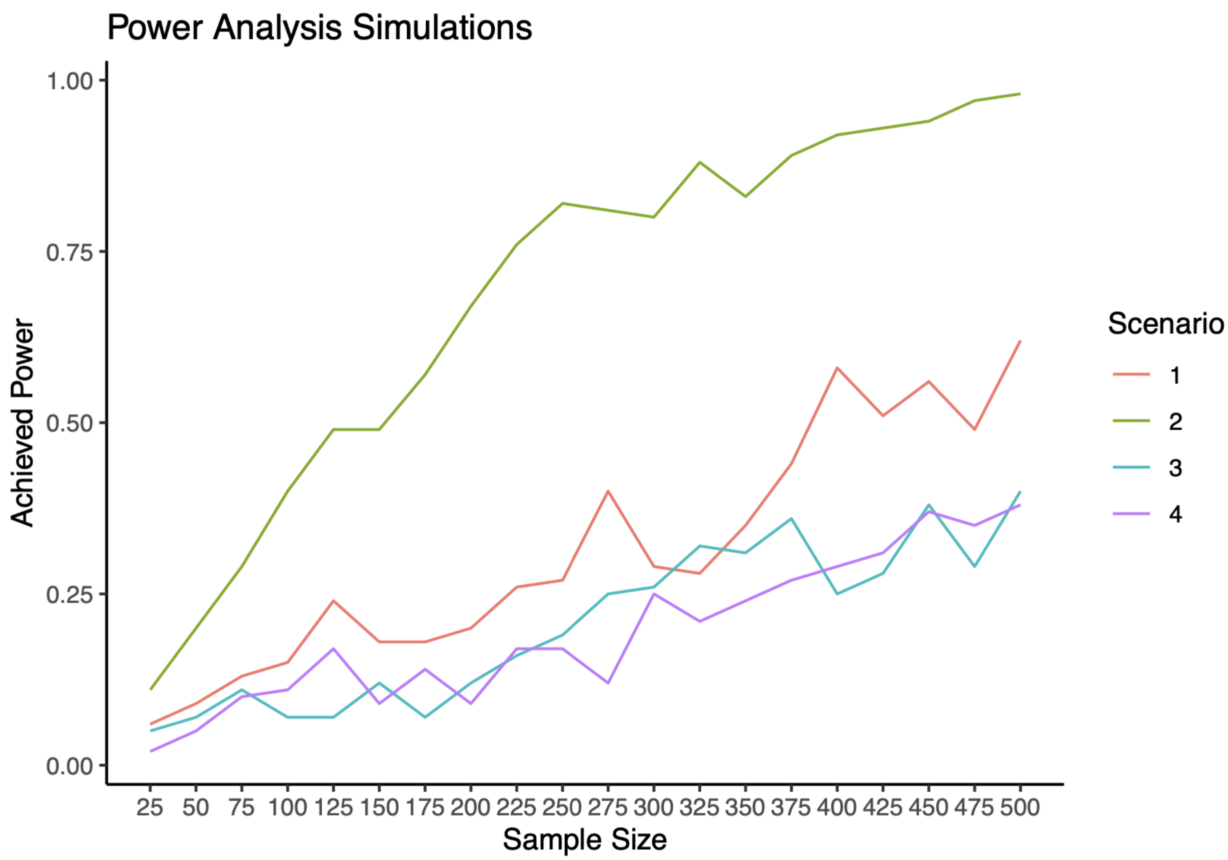

Figure 3: Power Analysis Simulations

As **Figure 3** shows, scenario 2 (blocking on party affiliation, with a uniform 4.3% effect size) shows the greatest level of power, followed by scenario 1. We deemed scenario 2 the most representative of our final experiment, as we would indeed be blocking by political party, and that the findings of Williamson et al.[11] serve as a reasonable basis for the effect we expect to see.

While we initially aimed for 250 participants to achieve sufficient statistical power (i.e., 80%), we successfully collected data from 200 participants (100 treatment, 100 control) in the final

[11] Williamson et al. (2021)

analysis. This difference in sample size is unlikely to dramatically change our findings, as the experiment's power was still high, but it may impact its significance.

Analysis

Data

The specific outcome being recorded from participants is their agreement on a 0-100 scale with each of the five pro-immigration survey statements. These statements are the same between the control and experimental groups. Several of the blocked variables in our randomization are recorded along with other preliminary self-reported information to help improve the efficiency of the analysis. The covariates and available options are the following:

Party: Republican, Independent, Democrat, Other
Political leaning: Strongly Conservative, Conservative, Lean Conservative, Moderate, Lean Liberal, Liberal, Strongly Liberal
Gender: Man, Women, Genderqueer, non-binary, questioning, prefer not to say, other
Age: 18-24, 25-34, 35-44, 45-54, 55-64, 65 or older, prefer not to disclose
Race or Ethnicity: Black or African American, American Indian or Alaska Native, White, Asian, Two or more, Native Hawaiian or Other Pacific Islander, Other race
Hispanic: Yes, No
Citizenship Status: U.S. citizen, Naturalized U.S. citizen, Lawful Permanent Resident, Not a U.S. citizen

Basic Model

In quantifying the effect of the treatment, the target variable was the average score across all five survey questions. This means the target variable still ranges from 0 - 100, where 0 is the lowest level of agreement, and 100 is the highest. The most basic causality assessment can be seen below in **Table 1, indicating** a positive but non-significant ATE (p-value = 0.304). Additionally, an R-squared value of < .01 was recorded, indicating the impacts of noise in only assessing the singular binary treatment variable.

	coef	std err	t	P>|t|	[0.025	0.975]
const	51.0120	3.196	15.961	0.000	44.709	57.315
is_treatment	4.7560	4.615	1.031	0.304	−4.344	13.856

Table 1: Causality Assessment Showing Positive But Non-Significant ATE

Improved Model

Covariate analysis suggests that efficiency gains were bound to be discovered by including them in the regression analysis. In **Figures 4** and **5,** density plots of average agreement scores in control and treatment can be seen across party labels. This demonstrates the potential significance of discerning baseline agreement levels and the clear divide in outcomes recorded.

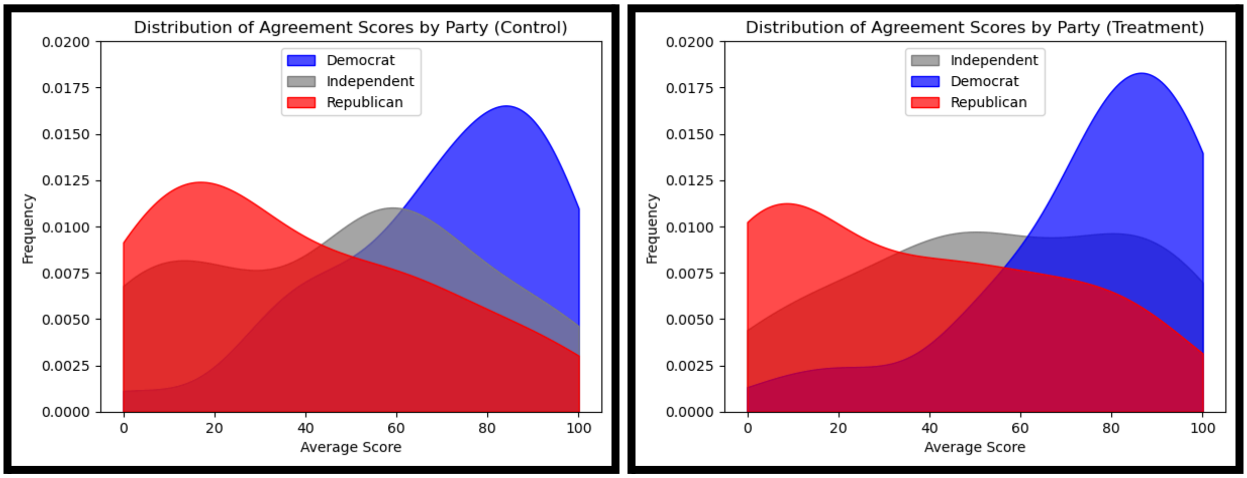

Figure 4 & Figure 5: Distribution of Agreement Scores by Party

Including all recorded covariates led to a rise in efficiency, improving the R-squared value to 0.411 and a significant likelihood ratio test (p-value < .001) compared to the basic model. As seen in **Table 2** below, political leaning and party affiliation covariates show significant baseline coefficients different from zero. However, despite efficiency gains, the positive ATE is still non-significant (p-value = 0.196).

	coef	std err	t	P>|t|	[0.025	0.975]
const	84.3415	12.487	6.754	0.000	59.695	108.988
is_treatment	5.5953	4.331	1.292	0.198	-2.954	14.144
party_Independent	-16.0286	6.534	-2.453	0.015	-28.925	-3.133
party_Other	-19.2193	416.143	-0.046	0.963	-840.590	802.151
party_Republican	-19.6130	7.145	-2.745	0.007	-33.716	-5.510
leaning_Conservative	-46.8587	8.638	-5.425	0.000	-63.908	-29.809
leaning_Lean Conservative	-35.2067	11.312	-3.112	0.002	-57.534	-12.880
leaning_Lean Liberal	-17.3658	8.861	-1.960	0.052	-34.856	0.124
leaning_Liberal	-13.3913	7.745	-1.729	0.086	-28.678	1.896
leaning_Moderate	-17.8222	6.654	-2.679	0.008	-30.955	-4.690
leaning_Strongly Conservative	-34.7083	10.713	-3.240	0.001	-55.853	-13.564
gender_woman	-3.4672	4.584	-0.756	0.450	-12.515	5.581
age_25-34	7.2072	8.758	0.823	0.412	-10.080	24.494
age_35-44	2.0478	9.013	0.227	0.821	-15.741	19.837
age_45-54	2.6571	9.503	0.280	0.780	-16.101	21.415
age_55-64	0.5171	9.592	0.054	0.957	-18.415	19.449
age_65 or older	-6.6510	9.006	-0.738	0.461	-24.427	11.125
race_American Indian or Alaska Native	4.8878	14.937	0.327	0.744	-24.595	34.371
race_Asian	4.0313	9.960	0.405	0.686	-15.627	23.689
race_Black or African American	0.0547	7.661	0.007	0.994	-15.067	15.176
race_Native Hawaiian or Other Pacific Islander	11.0322	176.154	0.063	0.950	-336.656	358.721
race_Other race	5.4656	9.934	0.550	0.583	-14.141	25.073
race_Two or more	4.7455	10.526	0.451	0.653	-16.031	25.522
hispanic_Yes	-1.7514	8.275	-0.212	0.833	-18.085	14.582
citizenship_Lawful Permanent Resident	27.6849	18.802	1.472	0.143	-9.426	64.796
citizenship_Naturalized U.S. citizen	-12.2263	11.244	-1.087	0.278	-34.419	9.966
citizenship_Not a U.S. citizen	18.4289	76.200	0.242	0.809	-131.973	168.831

Table 2: Causality Analysis With All Covariates

Final Modeling

	coef	std err	t	P>|t|	[0.025	0.975]
Democrat_x_treatment	6.1832	5.588	1.106	0.270	−4.848	17.214
Republican_x_treatment	−1.8775	6.421	−0.292	0.770	−14.553	10.798
Independent_x_treatment	5.0448	7.891	0.639	0.523	−10.532	20.622
Other_x_treatment	−6.7124	5.663	−1.185	0.238	−17.891	4.466

Table 3: Analysis of Heterogeneous Treatment Effects

Heterogeneous treatment effects were considered throughout the analysis process. Several covariates were tested by crossing the different statuses with the treatment variable. In all cases, the likelihood ratio test failed, denoting that the inclusion of the crossed variables did not add additional predictable power to the model. **Table 3** shows an example of crossing party identification with the treatment. Here, it's interesting to note that the magnitude of different political parties differs; however, all coefficients are not significant.

Discussion

After the iterative modeling approach, we found that the best R-squared value achieved was 0.411, and the treatment effect was not statistically significant. This less-than-optimal R-squared value underscores the amount of noise present in the data. It suggests that additional covariates may exist that could have reduced our standard errors to evaluate the effect more precisely. Other covariates that could be interesting to explore in future analysis are income, education level, and political engagement, all of which are important factors in political identity and potentially in stances on immigration.

Previous research studies moderated political persuasion and conformity through more well-known political figures or party messaging. Our experiment, focusing on simulated peers, suggests that peer conformity regarding political beliefs may be less effective than that of political leaders. Additionally, ATEs were typically much smaller in digital studies than the results recorded in this experiment; however, significantly larger sample sizes in those experiments helped capture significant effects.

Conclusion

While previous political research suggests that messages from elites and well-recognized political sources can sway public opinion on immigration issues, a more general peer-level conformity paradigm may not behave similarly. Additionally, this simple experiment explores the intersection of political beliefs and social media, which has become a focal point of the past few election cycles. In the age of increasing political polarization, is this exacerbated by political messaging on social media? As discussed above, while this research did not uncover significant heterogeneous treatment effects by party, could this be possible in a social media context, and are these heterogeneous effects inherently different from other mediums of political discourse? This then begs the question: what is ethical moderation in a political social media space that may hold the power of persuasion of its users?

Works Cited

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. Science, 337(6092), 337-341.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295-298.

Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. Proceedings of the National Academy of Sciences, 107(12), 5334-5338.

Gaucher, D., Friesen, J., Neufeld, K., & Esses, V. (2018). Changes in the Positivity of Migrant Stereotype Content: How System-Sanctioned Pro-Migrant Ideology Can Affect Public Opinions of Migrants. Social Psychological and Personality Science, 9(2), 223-233.

PureSpectrum Marketplace. (n.d.). Retrieved from www.purespectrum.com.

Qualtrics. (n.d.). Retrieved from www.qualtrics.com.

United States Census Bureau. (2020). DP1 | Profile of General Population and Housing Characteristics. Retrieved from https://data.census.gov/table?g=010XX00US&d=DEC+Demographic+Profile.

Williamson, S., Adida, C., Lo, A., Platas, M., Prather, L., & Werfel, S. (2021). Family Matters: How Immigrant Histories Can Promote Inclusion. American Political Science Review, 115(2), 686-693.

Zaller, J. R. (1992). The nature and origins of mass opinion. Cambridge University Press.

Zhirkov, K., Verkuyten, M., & Ponarin, E. (2021). Social conformity and prejudice toward immigrants: the role of political messaging. Social Influence, 16(1), 65-77.