

Boese-Schlosser, Vanessa; Klatt, Nikolina; Meißner, Daniel; Ngiam, Janice; Ziblatt, Daniel

Working Paper

How survey design influences people's understanding of democracy

WZB Discussion Paper, No. SP V 2025-504

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Boese-Schlosser, Vanessa; Klatt, Nikolina; Meißner, Daniel; Ngiam, Janice; Ziblatt, Daniel (2025) : How survey design influences people's understanding of democracy, WZB Discussion Paper, No. SP V 2025-504, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:

<https://hdl.handle.net/10419/328005>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Vanessa Boese-Schlosser

Nikolina Klatt

Daniel Meißner

Janice Ngiam

Daniel Ziblatt

How Survey Design Influences People's Understanding of Democracy

Discussion Paper

SP V 2025-504

September 2025

Research Area

Dynamics of Political Systems

Research Unit

Transformations of Democracy

WZB Berlin Social Science Center
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Copyright remains with the author(s).

Discussion papers of the WZB serve to disseminate the research results of work in progress to encourage the exchange of ideas and academic debate. Inclusion of a paper in the discussion paper series does not constitute publication and should not limit publication in any other venue. The discussion papers published by the WZB represent the views of the respective author(s) and not of the institute as a whole.

Vanessa Boese-Schlosser, Nikolina Klatt, Daniel Meißner, Janice Ngiam, Daniel Ziblatt

How Survey Design Influences People's Understanding of Democracy

SP V 2025-504

Wissenschaftszentrum Berlin für Sozialforschung (2025)

Affiliation of the authors:

Vanessa Boese-Schlosser

Berlin Social Science Centre (WZB)

Nikolina Klatt

Freie Universität Berlin & WZB Berlin Social Science Center

Daniel Meißner

Freie Universität Berlin & WZB Berlin Social Science Center

Janice Ngiam

Charite – Universitätsmedizin Berlin & WZB Berlin Social Science Center

Daniel Ziblatt

WZB Berlin Social Science Center & Harvard University

Abstract

While it is well established that closed-ended survey questions can prime respondents, less is known about how such priming affects the content of subsequent open-ended responses. We investigate this effect in the context of different understandings/perceptions of the term democracy using a survey experiment conducted in Germany in December 2023. Respondents ($N = 3,123$) were randomly assigned to a control or treatment group: the control group answered an open-ended question about what democracy meant to them first, while the treatment group first responded to closed-ended questions about democratic decision-making before also answering the open-ended question. We then compare the open-ended responses across groups.

Using computational text analysis, we test three hypotheses regarding priming effects on (1) response length, (2) lexical overlap with vocabulary from the closed-ended question, and (3) thematic emphasis. Our findings indicate that while response length merely differs slightly between groups, primed participants are more likely to incorporate vocabulary from the closed-ended questions and to emphasize topics related to democratic decision-making. These results highlight how subtle differences in question order can meaningfully influence the way individuals express their understanding of democracy.

Keywords: Priming, computational text analysis, open-ended question, closed-ended questions, democracy understandings

How Survey Design Influences People's Understanding of Democracy

Vanessa A. Boese-Schlosser* Nikolina Klatt[†] Daniel Meißner[‡]
Janice Ngiam[§] Daniel Ziblatt[¶]

September 25, 2025

Abstract

While it is well established that closed-ended survey questions can prime respondents, less is known about how such priming affects the content of subsequent open-ended responses. We investigate this effect in the context of different understandings/perceptions of the term *democracy* using a survey experiment conducted in Germany in December 2023. Respondents ($N = 3,123$) were randomly assigned to a control or treatment group: the control group answered an open-ended question about what democracy meant to them first, while the treatment group first responded to closed-ended questions about democratic decision-making before also answering the open-ended question. We then compare the open-ended responses across groups.

Using computational text analysis, we test three hypotheses regarding priming effects on (1) response length, (2) lexical overlap with vocabulary from the closed-ended question, and (3) thematic emphasis. Our findings indicate that while response length merely differs slightly between groups, primed participants are more likely to incorporate vocabulary from the closed-ended questions and to emphasize topics related to democratic decision-making. These results highlight how subtle differences in question order can meaningfully influence the way individuals express their understanding of democracy.

Keywords: Priming, computational text analysis, open-ended question, closed-ended questions, democracy understandings

*WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany.

[†]Freie Universität Berlin & WZB Berlin Social Science Center

[‡]Freie Universität Berlin & WZB Berlin Social Science Center

[§]Charité – Universitätsmedizin Berlin & WZB Berlin Social Science Center

[¶]WZB Berlin Social Science Center & Harvard University

1 Introduction

In recent years, the global landscape has witnessed a wave of autocratization, with democratic institutions being eroded not only in fragile democracies but also in long-established ones (Boese et al., 2022b; Levitsky and Ziblatt, 2018; Lührmann and Lindberg, 2019). Populist leaders have increasingly portrayed themselves as defenders of the people while simultaneously undermining core democratic norms and checks and balances (Bermeo, 2016; Waldner and Lust, 2018). Against this backdrop, it is more important than ever to understand how ordinary citizens conceptualize democracy: What comes to mind when they hear the term, and what do they believe constitutes democratic governance?

Existing survey-based research on citizens' understandings of democracy has largely relied on closed-ended items that offer a set of predefined response options (e.g. Ferrín and Kriesi, 2016; Landwehr and Steiner, 2017; Wunsch et al., 2025). While such approaches facilitate comparability and structured analyses, they may a) constrain respondents' ability to express more idiosyncratic views and b) introduce priming effects by making the aspects of democracy in the response options more cognitively accessible than they would have been spontaneously. In this sense, closed-ended formats risk shaping rather than merely capturing public opinion—particularly on complex and contested concepts like democracy.

Another context in which priming effects may arise is when open-ended questions follow batteries of closed-ended items within the same survey. This format is common in large-scale surveys such as the European Social Survey (ESS) or the GESIS Panel, which include hundreds of primarily closed-ended questions and a small number of open-ended prompts (e.g. “other, specify”) . These responses are often treated as windows into respondents' “unfiltered” thinking, offering insights unconstrained by predefined categories. Our study raises a critical methodological concern: the order in which these questions are presented may systematically bias the content of such open-ended responses.

We thus explore the research question:

Does prior exposure to closed-ended questions about democratic decision-making influence the way respondents answer subsequent open-ended questions about democracy, and if so how?¹

We test three hypotheses on potential priming effects: (1) that exposure to closed-ended questions increases the length of open-ended responses; (2) that it leads participants to echo the vocabulary of earlier items; and (3) that it shifts the thematic focus toward dimensions emphasized in the closed-ended format.

We address these questions using a survey experiment conducted in Germany in December 2023 with 3,123 participants. Respondents were randomly assigned to either a treatment group,

¹The hypotheses tested in this paper are part of a broader research project that was pre-registered on the Open Science Framework (OSF) prior to data collection. The pre-analysis plan is available at: <https://osf.io/a794r>. The present study focuses on a answering research question 4.2 (a) of the pre-analysis plan, i.e. “Is there a priming effect of closed-ended questions for participants' understandings of democracy?”. Appendix A describes deviations from the pre-analysis plan.

which received a battery of closed-ended questions about democratic decision-making before answering an open-ended question on democracy, or a control group, which answered the open-ended question first. The closed-ended items focused on one specific dimension of democracy (democratic decision-making) allowing us to test whether this priming increases attention to that dimension in the subsequent open-ended responses.

Results suggests that while the length of open-ended answers increases only slightly through priming, the content of open-ended responses differs substantially between treatment and control groups. Using structural topic modeling and keyword-based measures, we find that respondents exposed to closed-ended items about democratic decision-making are more likely to emphasize that dimension and use the corresponding vocabulary in their open-ended reflections. This suggests a priming effect, where the earlier questions shift respondents' focus even when later asked to express their views freely.

These findings carry important implications for the interpretation of open-ended survey data. Researchers often treat such responses as spontaneous expressions. Our study challenges that assumption. Even when respondents write freely, their answers may be shaped by the content and framing of preceding questions. As open-ended questions gain popularity in large-scale surveys and computational methods become standard tools for analyzing them, attention to priming effects becomes essential for valid inference.

2 Literature Review

Psychological research has shown that semantic priming — the influence of prior stimuli on subsequent responses — is a robust and well-documented phenomenon (Hutchison, 2003; Lucas, 2000). In survey research, earlier questions can shape how participants interpret and answer later ones. For example, Bullock and Rader (2022) show that the format of political knowledge questions significantly affects reported levels of knowledge. This raises the concern that responses to closed-ended questions about democracy understandings may reflect questionnaire design rather than pre-existing views.

A large body of empirical research explores how citizens understand democracy using closed-ended survey responses (Cho, 2014; Gessler and Wunsch, 2025; Landwehr and Steiner, 2017; Wunsch et al., 2025). Two recent studies provide comprehensive overviews of this survey based “understandings of democracy” literature, highlighting issues related to conceptualization, operationalization, measurement, aggregation, and social desirability biases (König et al., 2022; Osterberg-Kaufmann et al., 2020). Yet, potential priming effects through survey design features, such as question order, are outside the scope of these articles suggesting that a potential key dimension of measurement bias remains underexplored.

Several studies use qualitative methods, such as interviews, focus groups, or ethnographic observation, to explore how citizens understand democracy in more open-ended and discursive ways (e.g. Espinosa et al., 2024; Lu and Shi, 2015; Schäfer et al., 2025). Some even combine these with survey instruments (e.g. Wunsch, 2025). These approaches offer valuable depth and are well-suited for capturing how individuals articulate and refine their views through

dialogue. At the same time, large-scale surveys allow researchers to reach broader populations in a consistent manner across contexts, making them indispensable for comparative research. Rather than being alternatives, qualitative and quantitative approaches are complementary, each illuminating different facets of how people understand democracy. Yet, within survey-based research, especially when open-ended responses follow closed-ended questions, the possibility that survey design itself may shape responses remains underexplored.

The survey based research on understandings of democracy provides an ideal case for studying priming effects because democracy is an inherently multidimensional concept (Boese et al., 2022a; Teorell et al., 2019). Even within the political science literature, dimensions such as political participation, liberal rights, rule of law, and decision-making coexist and are emphasized to varying degrees depending on the definition and the research question at hand (Boese, 2019; Coppedge et al., 2011; Fjelde et al., 2021). We take advantage of this multidimensionality: In the closed-ended items, we focus on the dimension of democratic decision-making. By priming only this specific aspect, we can test whether respondents are subsequently more likely to invoke it in their open-ended responses, compared to other democratic dimensions that were not emphasized beforehand.

Structural Topic Modeling (STM) is a machine-learning-based approach designed to identify latent thematic structures in textual data allowing researchers to assess how these vary across individual- or group-level covariates (Roberts et al., 2014). While STM has been widely discussed in methodological literature (e.g. Weston et al., 2023; Ying et al., 2022), substantive applications in the social sciences remain relatively limited (e.g. Bagozzi and Berliner, 2018; Blaydes et al., 2018), and its potential to examine effects of survey design choices, such as question order, has not been systematically explored.

Our study contributes to this emerging field by combining experimental variation in question order with STM-based analysis of open-ended responses. Specifically, we assess whether closed-ended questions prime respondents to write longer answers, adopt specific vocabulary, or emphasize different themes when describing democracy. In doing so, we highlight how STM can be leveraged not only to uncover thematic content but also to detect subtle framing effects embedded in survey design.

3 Data & Approach

We conducted a survey experiment to investigate how exposure to closed-ended questions influences participants' open-ended responses about democracy. The survey experiment was fielded by the survey company Bilendi in Germany between December 20th-28th, 2023. The questionnaire was administered in German. Specifically, we examine whether the order in which questions are presented — open-ended or closed-ended first — affects the length and/or the content of open-ended responses.

The participants were randomly assigned to either a control or a treatment group. A comprehensive balance check indicates only minor imbalances between treatment and control group with regard to observable characteristics. Table 7 in D provides a complete overview of

the balance check by treatment status. Robustness checks that included covariates exhibiting minor imbalances based on standardized mean differences (SMD) yielded consistent results. The control group received the open-ended question first, prompting them to reflect freely on the concept of democracy without any preceding cues. The treatment group, by contrast, was first exposed to a battery of closed-ended questions about democratic decision making before answering the same open-ended prompt. This setup allows us to assess how exposure to information about democratic decision making influences the way individuals later formulate their own views on what democracy entails.

Our dataset includes responses from 3,123 participants, with 1,595 assigned to the control group and 1,528 to the treatment group.

To prepare the data for analysis, we removed incomplete entries, corrected spelling errors, and computed two key outcome variables for each open-ended response: (1) the total word count (to evaluate Hypothesis 1), and (2) the percentage of thematic overlap with the vocabulary from the closed-ended questions (to evaluate Hypothesis 2).

Approach: To examine how the order of question types affects open-ended responses, we specified and tested the following hypotheses:

Hypothesis 1: The number of words written in the open-ended response will be larger if participants are first exposed to the closed-ended questions (treatment group) compared to those who receive the open-ended question first (control group).

Hypothesis 2: Participants will make greater use of vocabulary from the closed-ended questions in their open-ended responses if they were first exposed to the closed-ended questions.

Hypothesis 3: The thematic content of open-ended responses will differ systematically depending on whether participants were first exposed to closed-ended questions. Specifically, participants in the treatment group will be more likely to emphasize themes introduced in the closed-ended questions, whereas participants in the control group will draw on a broader or more spontaneous set of associations.

Hypothesis 1 posits that participants who were first exposed to closed-ended questions (treatment group) would provide longer open-ended responses than those who received the open-ended question first (control group). To test this, we computed the total word count of each open-ended response. After assessing the distribution of word counts using the Shapiro-Wilk test, we found that normality assumptions were violated for both groups. As a result, we employed the non-parametric Wilcoxon rank-sum test to compare median word counts between groups. We also estimated a logistic regression model predicting treatment assignment based on word count, to further assess whether response length systematically varied by group.

Hypothesis 2 investigates whether treatment group participants incorporated more words from the closed-ended questions into their open-ended responses. To quantify this, we calculated the percentage of lexical overlap between each participant's open-ended response and the vocabulary set drawn from the closed-ended question items. As with word counts, we first

conducted Shapiro-Wilk tests, which indicated non-normal distributions of the overlap measure in both groups. We again relied on a Wilcoxon rank-sum test to compare the two distributions. Additionally, we estimated a logistic regression model predicting treatment assignment based on the overlap metric. In addition, we conducted a targeted analysis of the word stem “*entscheid*” (decision), which appeared in 12 of 18 closed-ended questions, to assess whether this word subsequently was used more often in the treated participants’ open-ended responses.

Hypothesis 3 assesses whether the treatment influences the substantive content of participants’ open-ended responses. To formally test Hypothesis 3, we applied a Structural Topic Model (STM) to the open-ended responses (with unlimited iterations) and modeled the relationship between document-level covariates and topic prevalence.² In the context of STM, a *topic* refers to a cluster of words that frequently occur and are interpreted as representing a coherent theme or subject. Each open-ended response was treated as one *document*, and the main covariate of interest was treatment assignment (i.e., whether a respondent was first exposed to the closed-ended items). The STM algorithm estimates the proportion of each topic in a response (document) as a mixture of topics. We then fit a regression of these estimated topic proportions on the treatment variable, using a logistic-normal regression model. The output reflects the marginal difference in expected topic proportions between the treatment and control groups, holding all else constant.

STM requires the number of topics K to be specified in advance. Yet, there is no universally accepted method for determining the optimal number of topics: One option is to use diagnostic metrics, another option is to let theoretical considerations and interpretability guide model selection.³ In the context of answering Hypothesis 3, our primary interest is assessing whether the treatment influenced the thematic content of responses in a systematic way, and whether there are any observable patterns in terms of topics covered - rather than the theoretical interpretation of each individual topic and determining an optimal number of topics. Hence, began our model selection using four established STM diagnostic metrics: held-out likelihood, residuals, semantic coherence, and the evidence lower bound (ELBO). A more detailed description and analysis of these metrics can be found in C. The model diagnostics then guided our selection from a broader ($K = 5 - 40$) to a narrower range of topic numbers ($K = 3 - 15$).⁴ Based on these evaluations, we focus on models with $K = 4, 6, 8, 10$, and 12 , which balance coherence, interpretability, fit, and parsimony. Below, we compare the results across this set of models.

²Prior to running the STM, we lemmatized and preprocessed all open-ended responses (including tokenization, lowercase, punctuation removal).

³Bonikowski and Nelson, 2022; Weston et al., 2023, provide comprehensive overviews of such model selection options.

⁴See Figures 9 and 10 in C.

Table 1: Logistic regression modeling the probability of membership in control/treatment group as a function of the total number of words

Logistic Regression Model			
Predictor	log(OR)	95% CI	p-value
Total Number of Words ^a	0.02	0.00, 0.03	0.012
Number of Observations	3,123		

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

^a This refers to the total number of words in a response.

4 Results

4.1 Hypothesis 1: Response Length

Hypothesis 1 proposed that participants in the treatment group—those who answered closed-ended questions before the open-ended prompt—would provide longer responses than participants in the control group. Figure 3 in B.1 displays the distribution of the total number of words written (wordcount) by the respondents for the control and the treatment group. Contrary to expectations, the difference in response length is extremely small. A Wilcoxon rank-sum test (data not normally distributed) indicated no significant difference in word count between the two groups ($p = 0.138$).

To further explore this relationship, we estimated a logistic regression model predicting treatment assignment based on the total number of words in the open-ended response. The results are displayed in Table 1. They show a small but statistically significant association ($\beta = 0.02$, $p = 0.012$), suggesting that longer responses were slightly more likely to come from the treatment group. In sum, we find modest support for H1: while the median word count does not differ significantly across groups, a logistic regression suggests that longer responses are slightly more likely in the treatment group.

4.2 Hypothesis 2: Lexical Overlap with Closed-Ended Vocabulary

Hypothesis 2 posited that participants in the treatment group would incorporate more vocabulary from the closed-ended questions in their open-ended responses. Figure 4 in B.2 shows the density plot of the percentage of lexical overlap between respondents answers and the words used in the closed-ended questions for control and treatment group. The results provide strong support for this hypothesis.

Participants in the treatment group exhibited significantly greater lexical overlap with the closed-ended question vocabulary compared to those in the control group ($p < 0.001$, Wilcoxon rank-sum test). A logistic regression model confirmed this result (see Table 2), showing a positive and significant relationship between lexical overlap and treatment assignment ($\beta = 0.01$, $p < 0.001$). These findings indicate that exposure to specific terms about democratic decision-

Table 2: Logistic regression modeling the probability of membership in control/treatment group as a function of the lexical overlap in percent.

Logistic Regression Model			
Predictor	log(OR)	95% CI	p-value
Lexical Overlap (in %) ^a	0.01	0.01, 0.02	<0.001
Number of Observations	3,123		

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

^a This refers to the lexical overlap as a percentage, i.e., the overlap of words used in the open-ended responses with words from the closed-ended survey items.

making prior to the open-ended question increased the likelihood that participants would echo that language in their responses.

Analyzing the lexical overlap in more detail, we examined the usage of words containing the stem “*entscheid*” (e.g., *entscheiden* - to decide, *Entscheidung* - decision). This stem was particularly relevant because it appeared in 12 out of the 18 closed-ended items, potentially priming participants in the treatment group. Non-parametric tests (Wilcoxon rank-sum and Kruskal–Wallis) revealed a highly significant difference between groups ($p \approx 1.15 \times 10^{-14}$), with treated participants using “*entscheid*”-related words substantially more often than control participants.

A logistic regression analysis confirmed this finding: each additional occurrence of a word containing the stem “*entscheid*” was associated with an 82% increase in the odds of belonging to the treatment group ($\ln(\text{OR}) = 0.60$, $\text{OR} = 1.82$, 95% CI [1.57, 2.12], $p < 0.001$). This analysis highlights how strongly the specific vocabulary in closed-ended questions can influence respondents’ subsequent open-ended responses.

4.3 Hypothesis 3: Differences in Topic Prevalence

Hypothesis 3 posited that participants exposed to the closed-ended questions first (treatment group) would emphasize different themes in their open-ended responses compared to participants in the control group.

We find substantial and statistically significant differences in topic prevalence between the treatment and control groups. Importantly, these differences persist across multiple model specifications. Because STM requires that the number of topics (K) be specified ex ante, we test the robustness of our findings by running models with varying numbers of topics.

We observe significant differences in topic prevalence (i.e. thematic content) regardless of how many number of topics we specify. This suggests that priming through closed-ended questions substantially shapes the content respondents articulate in open-ended formats. Figure 1 displays the persistent differences in topic prevalence between control and treatment effect

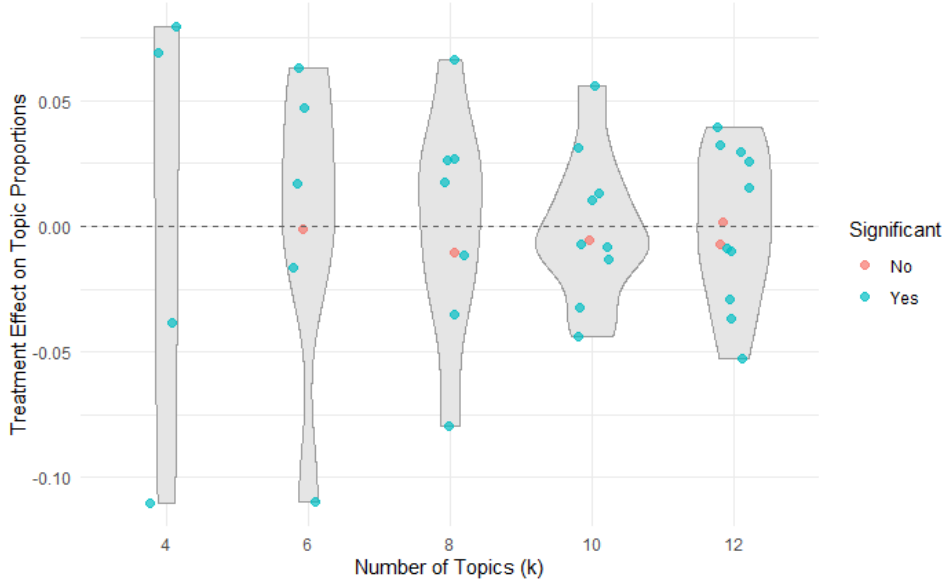


Figure 1: Distribution of treatment effects on topic proportions across STM models with varying numbers of topics (k). Each violin displays the distribution of estimated effects for all topics in a given STM solution (k). Individual points represent topic-specific effect estimates, with color indicating whether their 95% confidence intervals exclude zero (“significant effect”, $p < 0.05$). The dashed horizontal line denotes zero effect. This visualization summarizes the range and prevalence of treatment effects and highlights which topic effects are statistically significant for each value of k .

across different model specifications. Figure 2 illustrates the treatment effects by topic shedding light on the patterns of differences in thematic content between treatment and control group.

Figure 1 illustrates the estimated treatment effects on topic proportions for models with varying numbers of topics, where each point corresponds to a single topic and effects are grouped by K . Except for a few topics highlighted in red, the majority exhibit significant treatment effects, indicating consistent differences in topic prevalence between treatment and control groups. Notably, effect sizes tend to diminish as the number of topics increases, likely reflecting decreased semantic coherence and greater topic fragmentation. As K grows, topics become more granular and less distinguishable, reinforcing the earlier rationale to focus hypothesis testing on models with more coherent and well-defined topics.

To illustrate how the thematic content/topics covered by respondents in control and treatment group differed, Figure 2 displays the estimated treatment effects by topic, ordered from the largest negative effects at the top to the largest positive effects at the bottom.

Among the topics exhibiting strong positive treatment effects – i.e., those topics that were more prevalent in the treatment group – several prominently feature words containing the stem “entscheid” (e.g., *entscheidung*, *entscheiden*). This highlights a pronounced emphasis on democratic decision-making within the treatment group. The repeated presence of “entscheid”-related vocabulary across multiple model specifications underscores the robustness of this finding. Words with the stem “entscheid” appear among the top words for topics $K4_T1$, $K8_T8$, $K6_T6$, $K12_T8$, and $K10_T5$, all of which are associated with large positive effects. This pattern provides compelling evidence that the treatment effectively shaped participants’ responses to

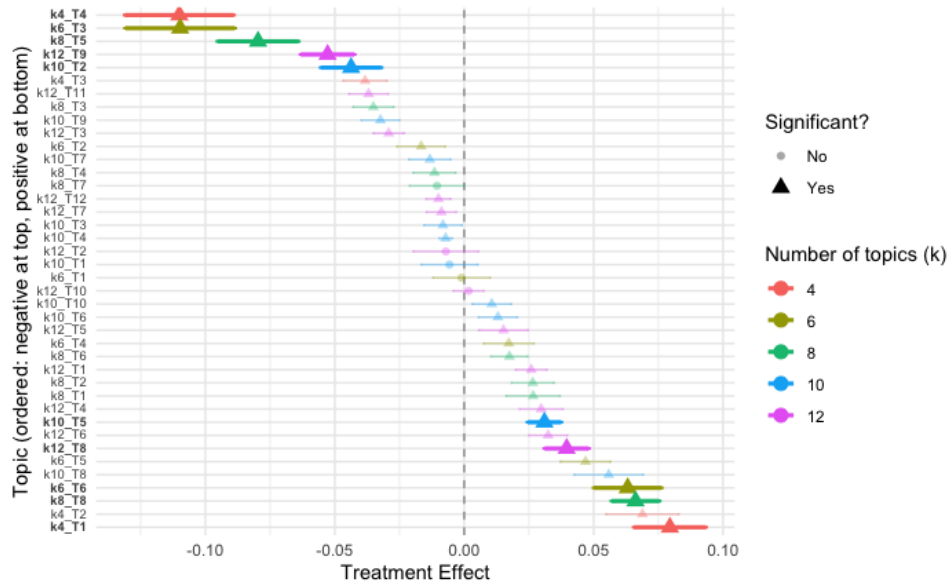


Figure 2: Caterpillar plot of estimated treatment effects on topic proportions from STM models with varying numbers of topics (k). Each point represents the estimated effect for a single topic, ordered by effect size across all models. Horizontal error bars denote 95% confidence intervals. Colors indicate assignment to a specific value of k , and shape distinguishes statistically significant effects (those for which the 95% confidence interval excludes zero, $p < 0.05$). The vertical dashed line marks zero effect. Topics highlighted by bold points and lines are those discussed in the text and summarized in Table 4 in B.3. This plot enables direct comparison of the magnitude, uncertainty, and statistical significance of topic-level treatment effects across STM solutions with different topic numbers.

focus on democratic decision-making processes, as this theme appears in two thirds of closed-ended survey items.

In contrast, topics with large negative treatment effects – i.e., those topics which were more prevalent in the control group – are characterized by frequent use of words containing the stem “*frei*” (free), reflecting a stronger focus on fundamental rights and civil liberties consistent across different model specifications. Words with the stem “*frei*” appear among the top words for topics *K6.T3*, *K4.T4*, *K8.T5*, *K12.T9*, and *K10.T2*, all of which are associated with large negative effects. These include concepts such as freedom (*frei*, *freiheit*), freedom of expression (*meinungsfreiheit*, *meinungsäußerung*), the right to vote (*wahl*, *wahlrecht*), as well as freedoms related to press, religion, and movement. Relative to the control group, treatment participants devoted a smaller share of their responses to individual rights and civil liberties themes and a larger share to decision-making themes. This suggests different emphases in how each group described democracy.

Table 4 in B.3 provides a comprehensive overview of the top words for the models with large treatment effects and highlights the difference in the occurrence of words with the stem “*entscheid*” and “*frei*” as discussed here.

5 Discussion and limitations

Our results suggest that citizens' understandings of democracy are not necessarily stable, deeply held beliefs, but may instead be responsive to contextual cues, even within the same survey. Individuals' understandings of democracy might often be developed through interactive and situational processes, rather than being fixed in advance.

These findings have important implications for future research using surveys to study political concepts like democracy, in particular in terms of survey design transparency/scope of inference, the interpretation of open-ended responses, as well as the value of mixed method approaches: First, researchers should explicitly report how survey design features—including question order, structure, and framing—may shape respondent answers. Priming effects from closed-ended questions can meaningfully influence subsequent open-ended responses. Thus, researchers should avoid overgeneralizing results from open-ended survey responses without accounting for the specific survey context in which they were elicited. Second, the flexibility observed in respondents' use of language and thematic focus suggests that open-ended answers may reflect not only underlying beliefs but also recent contextual inputs. Researchers should therefore be cautious about treating open-ended responses as unbiased windows into prior beliefs. One option to better capture the discursive and responsive elements of how citizens understand democracy, is to combine quantitative surveys with qualitative methods such as interviews, focus groups, and ethnographic observation. (Wunsch, 2025) is a great example for such an approach that creatively integrates survey instruments with focus group discussions to combine insights.

One limitation of our study concerns variation in response motivation, particularly among less politically engaged participants. Prior research has shown that many people, especially those less politically engaged, may lack sufficient motivation in survey settings to engage in deep memory retrieval or reflection (Boudreau and Lupia, 2011; Prior and Lupia, 2008). Consequently, the open-ended responses we analyze may reflect differing levels of cognitive effort and political interest, potentially confounding observed treatment effects. However, Table 7 in D demonstrates that the control and treatment groups are balanced across a wide range of relevant covariates. Additionally, robustness checks for Hypothesis 3, reported in 6 and 7 (Appendix B.3), show consistent results even after adjusting for covariates that exhibited slight imbalance based on standardized mean differences (income, state, ethnicity, religion, and political media consumption).

6 Conclusion

This study examines whether and how prior exposure to closed-ended questions influences participants' responses to an open-ended prompt about the meaning of democracy. Drawing on a randomized survey experiment with 3,123 respondents fielded in Germany in December 2023, we find evidence that priming through closed-ended questions affects both the form and content of open-ended responses: participants in the treatment group wrote slightly longer

answers, were significantly more likely to echo specific vocabulary from the closed-ended items, and emphasized different thematic understandings of democracy compared to the control group.

These findings contribute to broader literature on survey design and perceptions of democracy conceptions. Most notably, our results show that even subtle shifts in question order can influence how respondents articulate their views and the substantive content they mention. This not only highlights the malleability of democratic language, but also complicates assumptions that open-ended responses are purely spontaneous reflections of internal belief systems. For survey research, the results offer a cautionary note: question order and structure can shape not only what people say, but how they conceptualize abstract political ideas.

References

- Bagozzi, B. E., & Berliner, D. (2018). The politics of scrutiny in human rights monitoring: Evidence from structural topic models of us state department human rights reports. *Political Science Research and Methods*, 6(4), 661–677.
- Bermeo, N. (2016). On democratic backsliding. *Journal of democracy*, 27(1), 5–19.
- Blaydes, L., Grimmer, J., & McQueen, A. (2018). Mirrors for princes and sultans: Advice on the art of governance in the medieval christian and islamic worlds. *The Journal of Politics*, 80(4), 1150–1167.
- Boese, V. A. (2019). How (not) to measure democracy. *International Area Studies Review*, 22(2), 95–127.
- Boese, V. A., Gates, S., Knutsen, C. H., Nygård, H. M., & Strand, H. (2022a). Patterns of democracy over space and time. *International Studies Quarterly*, 66(3), sqac041.
- Boese, V. A., Lundstedt, M., Morrison, K., Sato, Y., & Lindberg, S. I. (2022b). State of the world 2021: Autocratization changing its nature? *Democratization*, 29(6), 983–1013.
- Bonikowski, B., & Nelson, L. K. (2022). From ends to means: The promise of computational text analysis for theoretically driven sociological research. *Sociological Methods & Research*, 51(4), 1469–1483.
- Boudreau, C., & Lupia, A. (2011). Political knowledge. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 171–183). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921452.011>
- Bullock, J. G., & Rader, K. (2022). Response options and the measurement of political knowledge. *British Journal of Political Science*, 52(3), 1418–1427.
- Cho, Y. (2014). To know democracy is to love it: A cross-national analysis of democratic understanding and political support for democracy. *Political Research Quarterly*, 67(3), 478–488.
- Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P., et al. (2011). Conceptualizing and measuring democracy: A new approach. *Perspectives on politics*, 9(2), 247–267.
- Espinosa, A., Janos, E., Pacheco, M., Cisneros, R., Cueva, F., De Las Casas De La Lama, R., De La Cruz Bellido, C. N., Grados, C., Huapaya, C., Mercado, S., et al. (2024). A qualitative study on the social representations of populism and democracy in peru. *Frontiers in Political Science*, 6, 1306060.
- Ferrín, M., & Kriesi, H. (2016). *How Europeans view and evaluate democracy*. Oxford University Press.
- Fjelde, H., Knutsen, C. H., & Nygård, H. M. (2021). Which institutions matter? re-considering the democratic civil peace. *International Studies Quarterly*, 65(1), 223–237.
- Gessler, T., & Wunsch, N. (2025). A new regime divide? democratic backsliding, attitudes towards democracy and affective polarization. *European Journal of Political Research*.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? a microanalytic review. *Psychonomic bulletin & review*, 10, 785–813.

- König, P. D., Siewert, M. B., & Ackermann, K. (2022). Conceptualizing and Measuring Citizens' Preferences for Democracy: Taking Stock of Three Decades of Research in a Fragmented Field. *Comparative Political Studies*, 55(12), 2015–2049. <https://doi.org/10.1177/00104140211066213>
- Landwehr, C., & Steiner, N. D. (2017). Where democrats disagree: Citizens' normative conceptions of democracy. *Political Studies*, 65(4), 786–804.
- Levitsky, S., & Ziblatt, D. (2018). *How democracies die*. Crown Publishing Group.
- Lu, J., & Shi, T. (2015). The battle of ideas and discourses before democratic transition: Different democratic conceptions in authoritarian china. *International Political Science Review*, 36(1), 20–41.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic bulletin & review*, 7, 618–630.
- Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: What is new about it? *Democratization*, 26(7), 1095–1113.
- Osterberg-Kaufmann, N., Stark, T., & Mohamad-Klotzbach, C. (2020). Challenges in conceptualizing and measuring meanings and understandings of democracy. *Zeitschrift für Vergleichende Politikwissenschaft*, 14(4), 299–320.
- Prior, M., & Lupia, A. (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science*, 52(1), 169–183.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, 58(4), 1064–1082.
- Schäfer, S., Syam, M., & Gogali, L. (2025). Living together beyond liberal democracy: Examples of local decision-making and managing resource extractivism in indonesia. *Frontiers in Political Science*, 7, 1370828.
- Teorell, J., Coppedge, M., Lindberg, S., & Skaaning, S.-E. (2019). Measuring polyarchy across the globe, 1900–2017. *Studies in Comparative International Development*, 54, 71–95.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4–70.
- Waldner, D., & Lust, E. (2018). Unwelcome change: Coming to terms with democratic backsliding. *Annual Review of Political Science*, 21(1), 93–113.
- Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, 6(2), 25152459231160105.
- Wunsch, N. (2025). *Democratic commitment: Why citizens tolerate democratic backsliding*. Oxford University Press.
- Wunsch, N., Jacob, M. S., & Derksen, L. (2025). The demand side of democratic backsliding: How divergent understandings of democracy shape political choice. *British Journal of Political Science*, 55, e39.

Ying, L., Montgomery, J. M., & Stewart, B. M. (2022). Topics, concepts, and measurement: A crowdsourced procedure for validating topics as measures. *Political Analysis*, 30(4), 570–589.

Online Appendix for “How Survey Design Influences People’s Understanding of Democracy”

A	Deviations from the Pre-Anaylsis Plan	(i)
A.1	Clarifying and Expanding the Hypotheses	(i)
A.2	Omission of Factor Comparisons	(ii)
A.3	Summary of Deviations and Justifications	(ii)
B	Additional Results	(iii)
B.1	Hypothesis 1: Response Length	(iii)
B.2	Hypothesis 2: Lexical Overlap and with Closed-Ended Questions	(iv)
B.3	Hypothesis 3: Differences in Topic Prevalence	(vi)
C	Topic Number Selection and Model Diagnostics	(x)
D	Balance Check by Treatment Status	(xii)

A Deviations from the Pre-Analysis Plan

This project is part of a larger data collection effort on public understandings of democracy in Germany. Prior to data collection, we pre-registered our research design and hypotheses on the Open Science Framework (OSF). The full pre-analysis plan is available here: <https://osf.io/a794r>.

Given the scope of the broader project, we decided to focus this particular paper on one specific component: the priming experiment. That is, we examine whether exposure to closed-ended questions on democratic decision-making influences how participants respond to a subsequent open-ended question about the meaning of democracy.

This focus corresponds to **Research Question 4.2(a)** of the pre-analysis plan:

Is there a priming effect of closed-ended questions for participants' understandings of democracy?

The pre-analysis plan also included an associated hypothesis:

The number of words written in the open-ended responses will be larger if the respondents first were primed with a closed-ended question.

This hypothesis is retained as *Hypothesis 1* in the present paper.

A.1 Clarifying and Expanding the Hypotheses

While the pre-analysis plan only explicitly stated a hypothesis regarding response length, it also outlined planned analyses on *content-related priming effects*, including:

- Comparing the appearance of keywords from Landwehr and Steiner (2017) across treatment and control groups.
- Descriptive comparisons of latent content structures (e.g., factors) across groups.

These analyses were merely described in section “**16. Statistical models for methodology 2a.(2)**” of the pre-analysis plan, and we now formalize them as additional hypotheses to enhance clarity and analytical transparency:

Hypothesis 2: Participants in the treatment group will use more vocabulary that overlaps with the wording of the closed-ended questions (lexical overlap).

Hypothesis 3: Thematic content of open-ended responses will differ systematically between treatment and control groups.

Thus, these hypotheses are not entirely new but reflect a formalization of planned analyses already embedded in the original design.

A.2 Omission of Factor Comparisons

One deviation from the pre-analysis plan concerns the comparison of latent factors extracted from open-ended responses across treatment and control groups. In the original plan, this step was intended as part of the descriptive analysis of content variation. However, in this paper we do not perform factor comparisons, as these factors are derived and interpreted in a separate part of the larger project focused on typologies of democratic understanding.

Because this paper isolates the priming experiment as a standalone contribution, we opted not to carry out cross-group factor comparisons that rely on constructs developed elsewhere in the broader project (Vandenberg and Lance, 2000). Instead, we focus on lexical overlap and topic modeling (STM) to assess content variation.

A.3 Summary of Deviations and Justifications

- We maintain the original hypothesis on response length (H1).
- We explicitly state two additional hypotheses (H2 and H3) related to content and thematic emphasis, which were part of the original analytic plan but not pre-formulated as hypotheses.
- We omit the planned factor comparison analysis due to its reliance on constructs developed in a separate part of the larger project.

These adjustments are made to enhance the analytical clarity and coherence of this paper, while remaining consistent with the pre-registered goals and structure.

B Additional Results

B.1 Hypothesis 1: Response Length

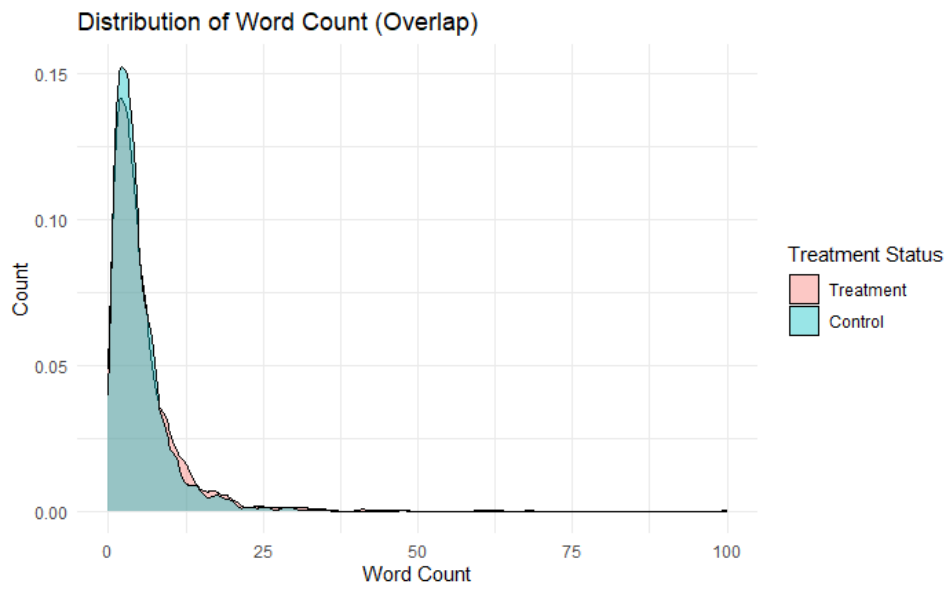


Figure 3: Distribution of word count across control and treatment group.

B.2 Hypothesis 2: Lexical Overlap and with Closed-Ended Questions

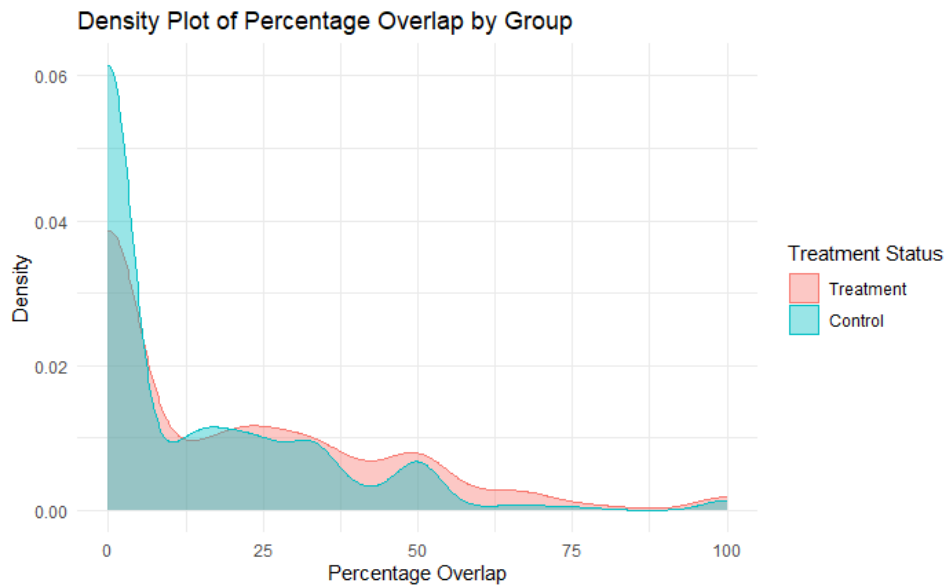


Figure 4: Lexical overlap of words used in the open-ended responses with words from the closed-ended survey items for participants in control and treatment groups

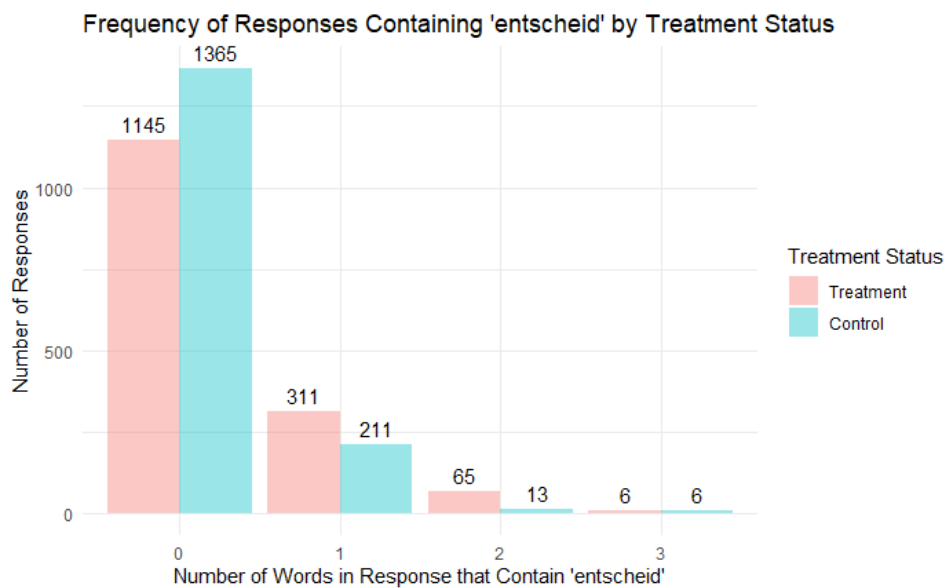


Figure 5: Comparison of the occurrence of the words containing “entscheid” across control and treatment group. Note: To improve clarity, responses with a count of four were excluded from the plot due to their very low frequency (only one observation).

Table 3: Logistic regression modeling the probability of membership in control/treatment group as a function of the number of "entscheid" words in a response.

Logistic Regression Model			
Predictor	log(OR)	95% CI	p-value
Number of "entscheid" words	0.60	0.45, 0.75	<0.001
Number of Observations	3,123		

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

B.3 Hypothesis 3: Differences in Topic Prevalence

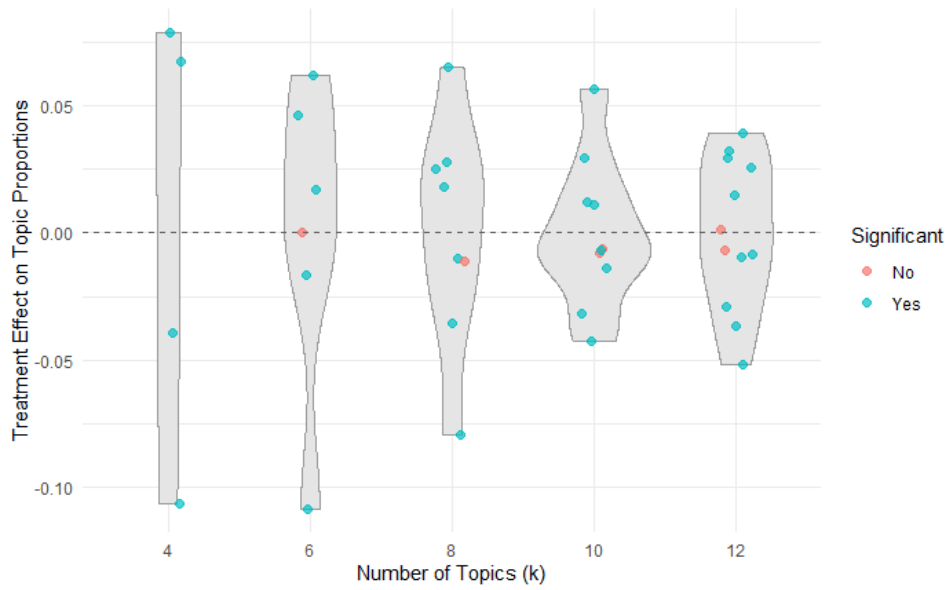


Figure 6: *Robustness check:* Distribution of treatment effects on topic proportions across STM models with varying numbers of topics (k) when additional covariates are included. Included covariates are income, state, ethnicity, religion and consumption of political media, which are all factor variables. Each violin displays the distribution of estimated effects for all topics in a given STM solution (k). Individual points represent topic-specific effect estimates, with color indicating whether their 95% confidence intervals exclude zero (“significant effect”, $p < 0.05$). The dashed horizontal line denotes zero effect. This visualization summarizes the range and prevalence of treatment effects and highlights which topic effects are statistically significant for each value of k .

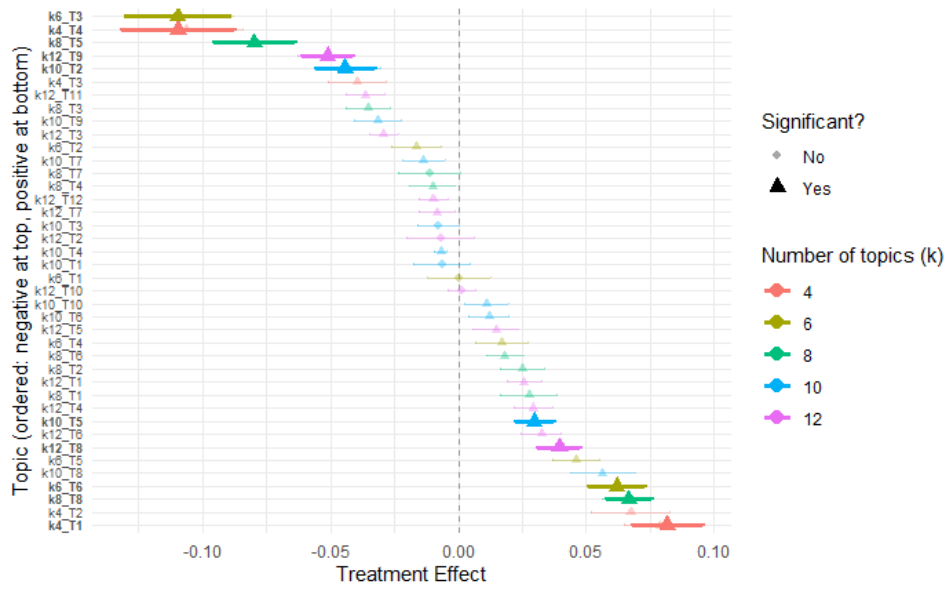


Figure 7: Robustness check: Caterpillar plot of estimated treatment effects on topic proportions from STM models with varying numbers of topics (k) when additional covariates are included. Included covariates are income, state, ethnicity, religion and consumption of political media, which are all factor variables. Each point represents the estimated effect for a single topic, ordered by effect size across all models. Horizontal error bars denote 95% confidence intervals. Colors indicate assignment to a specific value of k , and shape distinguishes statistically significant effects (those for which the 95% confidence interval excludes zero, $p < 0.05$). The vertical dashed line marks zero effect. Topics highlighted by bold points and lines are those discussed in the text and summarized in Table 4 in B.3. This plot enables direct comparison of the magnitude, uncertainty, and statistical significance of topic-level treatment effects across STM solutions with different topic numbers.

Topic ID	Metric	Top Words
Selected topics with large positive treatment effects:		
K4.T1 +0.079 (0.0071)	Prob	entscheidung, entscheiden, mehrheit, bevölkerung, vertret, wichtig, abstimmung
	FREX	mehrheit, bevölkerung, vertret, wichtig, abstimmung, gewählt, interess
	Lift	abschaffen, ausschließen, basieren, best, einhalten, eintreten, entscheidungsträg
	Score	mehrheit, entscheidung, bevölkerung, entscheiden, vertret, abstimmung, wichtig
K6.T6 +0.063 (0.0067)	Prob	demokrati, bedeuten, entscheiden, möglichkeit, gleich, kompromiss, wissen
	FREX	demokrati, bedeuten, kompromiss, wissen, heißen, schützen, leider
	Lift	austausch, berechtigten, berichten, bloß, demo, diskriminieren, durchsetzen
	Score	demokrati, bedeuten, entscheiden, heißen, gut, wissen, schützen
K8.T8 +0.066 (0.0047)	Prob	entscheidung, mehrheit, bevölkerung, treffen, minderheit, akzeptieren, mitbestimmen
	FREX	entscheidung, mehrheit, bevölkerung, treffen, minderheit, akzeptieren, mitbestimmen
	Lift	abwägen, allgemeinwohl, ausleben, beitrags, beteiligt, bewegung, bürgernah
	Score	entscheidung, mehrheit, treffen, bevölkerung, minderheit, mitbestimmen, akzeptieren
K10.T5 +0.031 (0.0033)	Prob	entscheidung, mitbestimmung, mehrheit, gleich, partei, treffen, land
	FREX	entscheidung, mehrheit, treffen, abgeordnet, mehrheitlich, regeln, einigkeit
	Lift	abwägen, beitrags, besetzen, bleiben, bürg, eintreten, entscheidend
	Score	mehrheit, entscheidung, mitbestimmung, treffen, partei, gleich, land
K12.T8 +0.040 (0.0043)	Prob	entscheidung, partei, land, wichtig, gehören, mehrheitsentscheidung, abstimmen
	FREX	partei, land, wichtig, mehrheitsentscheidung, gesellschaft, nehmen, grupp
	Lift	anerkennen, aug, ausgrenzung, beitrags, belang, bleiben, bundeskanz
	Score	entscheidung, land, wichtig, partei, gehören, abstimmen, abgeordnet
Selected topics with large negative treatment effects:		
K4.T4 -0.110 (0.0107)	Prob	frei, freiheit, wahl, meinungsfreiheit, meinungsäußerung, gerechtigkeit, gleichheit
	FREX	frei, freiheit, wahl, meinungsfreiheit, meinungsäußerung, gleichheit, wahlrecht
	Lift	bewegen, bewegungsfreiheit, eigentum, glaubensfreiheit, integr, mitgestaltung, repressali
	Score	frei, meinungsfreiheit, freiheit, wahl, meinungsäußerung, gleichheit, pressefreiheit
K6.T3 -0.110 (0.0108)	Prob	frei, freiheit, wahl, meinungsfreiheit, meinung, meinungsäußerung, gerechtigkeit
	FREX	frei, freiheit, wahl, meinungsfreiheit, meinungsäußerung, gerechtigkeit, gleichheit
	Lift	absicherung, absolut, achtung, annehmen, anonym, ausgewog, ausreis
	Score	frei, freiheit, meinungsfreiheit, wahl, meinungsäußerung, gleichheit, gerechtigkeit
K8.T5 -0.080 (0.0080)	Prob	frei, wahl, meinungsäußerung, wahlrecht, gewaltenteilung, pressefreiheit, unabhängig
	FREX	frei, wahl, meinungsäußerung, wahlrecht, gewaltenteilung, geheim, mitentscheiden
	Lift	ausreis, auswahl, bewegen, bewegungsfreiheit, bürgerbefragung, demonstrationsrecht, eigentum
	Score	frei, wahl, meinungsäußerung, gewaltenteilung, wahlrecht, unabhängig, geheim
K10.T2 -0.044 (0.0059)	Prob	frei, wahl, meinungsäußerung, gerechtigkeit, geheim, mitentscheiden, gemeinschaft
	FREX	frei, meinungsäußerung, geheim, demonstrationsrecht, legislativ, demonstrieren, exekutiv
	Lift	absolut, annehmen, bewegungsfreiheit, bürgerrechen, demonstrationsrecht, demonstrieren, einbinden
	Score	frei, wahl, meinungsäußerung, gerechtigkeit, geheim, demonstrationsrecht, mitentscheiden
K12.T9 -0.037 (0.0039)	Prob	frei, meinungsäußerung, pressefreiheit, religionsfreiheit, redefreiheit, parlament, reisefrei-
	FREX	frei, meinungsäußerung, pressefreiheit, religionsfreiheit, redefreiheit, reisefreiheit, demon-
	Lift	strationsrecht, annehmen, ausreis, bewegungsfreiheit, demonstrationsrecht, ein, einwohn, freie
	Score	frei, meinungsäußerung, pressefreiheit, reisefreiheit, demonstrationsrecht, redefreiheit, religionsfreiheit

Table 4: Key words for selected STM Topics with large treatment effects. Top words for each topic are shown based on four metrics: Highest Probability (Prob) (most frequent words within the topic), FREX (balancing frequency and exclusivity), Lift (words more unique to the topic relative to the corpus), and Score (a combined relevance measure). Effect size and direction below topic id, SE in brackets. Topics are ordered by number of topics K in the model. Words with stem "entscheid" are highlighted for topics with positive effects and words with stem "frei" are highlighted for topics with negative effects.

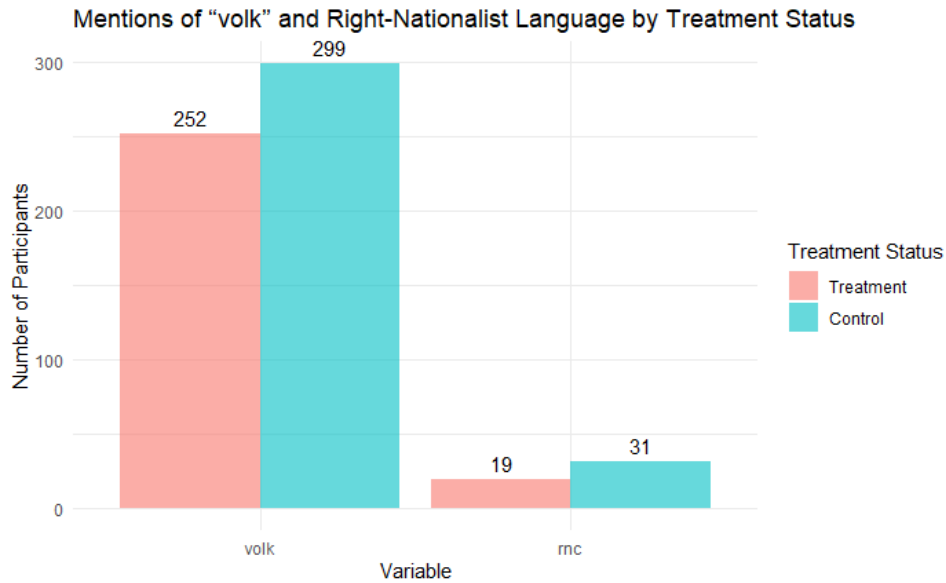


Figure 8: Bar plot showing the number of participants whose responses to the question “What is democracy to you?” included the term volk, as well as the subset of those responses that were manually coded as having a right nationalist connotation (rnc), grouped by treatment status. For each group (treatment and control), the plot displays the total number of participants mentioning volk and, within that, those whose response was judged to have a right nationalist meaning. Exact participant counts are shown above each bar. This visualization enables comparison of the prevalence of these themes across treatment status.

Table 5: Logistic regression modeling the probability of membership in control/treatment group as a function of the term “volk” occurring in a response.

Logistic Regression Model			
Predictor	log(OR)	95% CI	p-value
”volk” in response ^a	-0.16	-0.34, 0.03	0.10
Number of Observations	3,123		

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

^a Dummy variable = 1 if “volk” in response, else = 0.

Table 6: Logistic regression modeling the probability of membership in control/treatment group as a function of a response coded as having a right-nationalist connotation.

Logistic Regression Model			
Predictor	log(OR)	95% CI	p-value
Response with right-nationalist connotation ^a	-0.45	-1.0, 0.11	0.12
Number of Observations	3,123		

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

^a Dummy variable = 1 if response that included “volk” was manually coded as having a right nationalist connotation, else = 0.

C Topic Number Selection and Model Diagnostics

To determine the optimal number of topics for the structural topic model (STM), we compared models with a range of values for K by systematically evaluating them based on four standard STM diagnostic criteria or benchmarks to guide topic number selection: held-out likelihood, residuals, semantic coherence, and the lower bound (ELBO). Choosing a suitable topic number means balancing out the different metrics available. These values are visualized in Figure 9 for the range from 5 to 40 topics and in Figure 10 for the range from 3 to 15 topics. Based on the criteria, the broader range guided us towards the more narrow range of fewer topics.

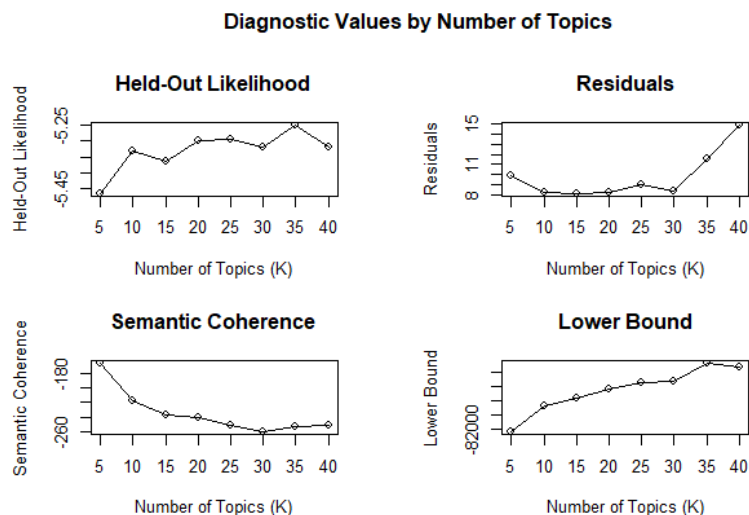


Figure 9: STM Diagnostic Values by Number of Topics (Unlimited iterations; $k = 5\text{--}40$)

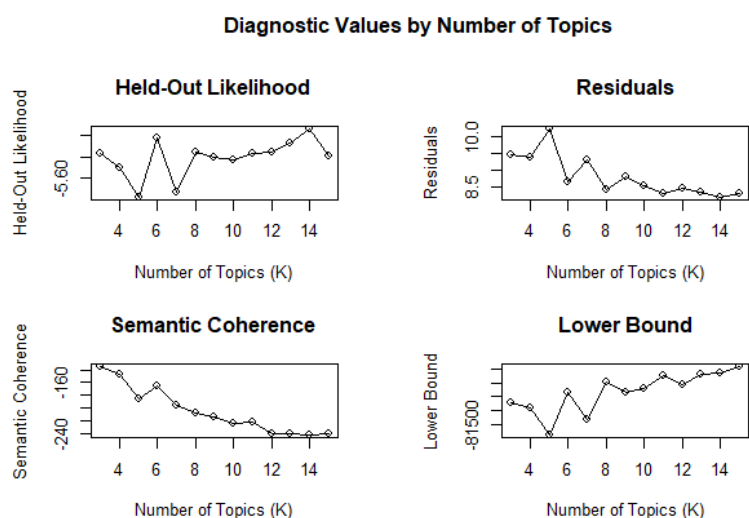


Figure 10: STM Diagnostic Values by Number of Topics (Unlimited iterations; $k = 3\text{--}15$)

- **Held-Out Likelihood:** This measure represents the per-document held-out log-likelihood. It is calculated based on the dataset in which 10% of documents have half of their words

removed. Based on the "held-out" words, it is measured how likely the trained model says the remaining text would be. Thus, a higher held-out likelihood means that a model's predictive ability is higher. Figure 9 shows that within the broader range of topics, held-out likelihood is highest for $K = 35$, however, the increase is marginal from $K = 10$, suggesting that additional topics added limited predictive power. Figure 10 shows that held-out likelihood across lower number of topics is highest for $K = 14$ but the difference is marginal when compared to most other specifications ($K = 3, 6, 8, 9, 10, 11, 12, 13$).

- **Residuals:** This measure calculates the degree of unexplained correlation between words after fitting the model, i.e. unexplained variation. A higher measure means that residuals are more dispersed and additional topics might reduce unexplained variance. Thus, lower residuals indicate higher explanatory power of the model. Figure 9 shows that residuals plateau between $K = 10$ and $K = 30$ and significantly increase for higher values. Figure 10 shows low residuals for $K \geq 10$.
- **Semantic Coherence:** This criterion measures semantic coherence based on the co-occurrence of top words in a topic. The more the most probable words in a topic are found together, the greater the semantic coherence. Mimno et al. (2021) demonstrate that this metric aligns closely with human assessments of topic quality. Figure 9 shows highest coherence for models with a lower number of topics ($K = 5, 10, 15$) and diminishing marginal decreases from there. A similar picture arises from Figure 10 where coherence is highest for models with lower number of topics and plateaus from $K = 12$. In general, higher K leads to less semantically coherent topics, which is consistent with topic fragmentation.
- **Lower Bound (ELBO):** This measure is an indicator of how well the model fits the observed data, it is an approximation of the log-likelihood. It also tracks model convergence. A model is considered converged when the bound only changes marginally between additional iterations. A less negative bound, i.e. a higher bound, indicates a better fit of the model. Figure 9 documents a gradually but diminishing marginal increase of the lower bound in the number of topics. Figure 10 shows that across models with lower number of topics, the increase of the lower bound is marginally decreasing from $K = 8$.

Taken together, these diagnostics, both from the bigger picture in Figure 9 and from the smaller range of models in Figure 10, suggest that looking at models with a lower number of topics $K = 4, 6, 8, 10, 12$ offers a good balance between interpretability, model fit, and parsimony. Therefore, we restrict the testing of differences in topic prevalence by treatment status (hypothesis 3) to these models.

D Balance Check by Treatment Status

This section provides a detailed balance check by treatment status of covariates that were included in the survey. Table 7 shows all continuous and categorical variables that were collected as part of the survey. These include sociodemographics such as age, gender, education, income, state, among others, and political information on the survey participants such as likelihood of voting for a specific party, participation in elections, self-reported interest in politics, consumption of political news, partisanship, among others. The sample was stratified based on the demographics age, gender, education, income, and state.

Table 7 displays means and standard deviations for continuous variables and counts and percentages for categorical variables. The results for testing differences between the treatment (CE first) and control (OE first) group show that randomization was generally successful. For continuous variables, a two-sample t-test (Welch's t-test by default) is used to compare the means between treatment and control groups. This test evaluates whether the group means are statistically significantly different, accounting for possible unequal variances. For categorical variables, a chi-square test (or Fisher's exact tests when counts are small) is used to compare the distribution of categories between treatment groups. The SMD measures the effect size or magnitude of difference between groups, independent of sample size. For continuous variables, it is calculated as the difference in group means divided by the pooled standard deviation. For categorical variables, it is computed from the difference in proportions standardized similarly. There is one significant difference detected at the 5%-level for religious affiliation (0.025). P-values for number of households members (0.085), self-employment (0.057), dual student (0.072), unemployed searching (0.097) and consumption of political news (0.054) are significant at the 10%-level. Values for the Standardized Mean Difference (SMD) are below 0.2 across all covariates and above 0.1 for income (0.110), state (0.108), ethnicity (0.104), religious affiliation (0.156) and consumption of political news (0.118), indicating only a minor imbalance between treatment and control group in observable characteristics.

Table 7: Balance check across covariates by treatment status

Variable/Category	Unit	Treatment Status		Tests for Difference	
		treatment	control	p-value	SMD
Number of observations	count	1528	1595		
Age	mean (SD)	48.02 (15.42)	47.26 (15.75)	0.177	0.048
Gender	count (%)			0.197	0.078
Female		751 (49.1)	824 (51.7)		
Male		775 (50.7)	765 (48.0)		
Non-binary		2 (0.1)	4 (0.3)		
Prefer not to disclose		0 (0.0)	2 (0.1)		
Education	count (%)			0.743	0.059
Apprenticeship		784 (51.3)	780 (48.9)		
Other degree		21 (1.4)	23 (1.4)		
Still in vocational training		45 (2.9)	43 (2.7)		
Technical school		205 (13.4)	239 (15.0)		
University/technical college degree		399 (26.1)	429 (26.9)		
Without qualification		74 (4.8)	81 (5.1)		
Income (in euros)	count (%)			0.149	0.110*

Table 7: Balance check across covariates by treatment status

Variable/Category	Unit	Treatment Status		Tests for Difference	
		treatment	control	p-value	SMD
0-499		50 (3.3)	52 (3.3)		
500-1,499		258 (16.9)	248 (15.5)		
1,500-2,499		365 (23.9)	403 (25.3)		
2,500-3,499		319 (20.9)	344 (21.6)		
3,500-4,499		262 (17.1)	222 (13.9)		
4,500-5,499		150 (9.8)	181 (11.3)		
5,500 or more		124 (8.1)	145 (9.1)		
State	count (%)			0.872	0.108*
Baden-Wuerttemberg		194 (12.7)	205 (12.9)		
Bavaria		222 (14.5)	245 (15.4)		
Berlin		59 (3.9)	64 (4.0)		
Brandenburg		46 (3.0)	48 (3.0)		
Bremen		14 (0.9)	19 (1.2)		
Hamburg		31 (2.0)	32 (2.0)		
Hesse		94 (6.2)	120 (7.5)		
Mecklenburg-Western Pomerania		27 (1.8)	31 (1.9)		
Lower Saxony		172 (11.3)	146 (9.2)		
North Rhine-Westphalia		341 (22.3)	357 (22.4)		
Rhineland-Palatinate		79 (5.2)	84 (5.3)		
Saarland		15 (1.0)	16 (1.0)		
Saxony		84 (5.5)	69 (4.3)		
Saxony-Anhalt		45 (2.9)	48 (3.0)		
Schleswig-Holstein		60 (3.9)	67 (4.2)		
Thuringia		45 (2.9)	44 (2.8)		
Country born in	count (%)			0.475	0.044
Federal Republic of Germany (FRG)		1157 (75.7)	1227 (76.9)		
German Democratic Republic (GDR)		286 (18.7)	294 (18.4)		
Other		85 (5.6)	74 (4.6)		
Marital status	count (%)				
Single		358 (23.4)	364 (22.8)		
In a relationship		316 (20.7)	348 (21.8)		
In a civil union		13 (0.9)	17 (1.1)		
Married		673 (44.0)	691 (43.3)		
Divorced		124 (8.1)	133 (8.3)		
Widowed		44 (2.9)	42 (2.6)		
Children	mean (SD)	0.57 (0.50)	0.56 (0.50)	0.550	0.021
Household status					
Number of household members	mean (SD)	2.29 (1.41)	2.44 (3.07)	0.085*	0.062
Number of children in household	mean (SD)	0.38 (0.76)	0.37 (0.79)	0.742	0.012
Employment status					
Self-employed	mean (SD)	0.04 (0.20)	0.06 (0.24)	0.057*	0.068
Employed	mean (SD)	0.60 (0.49)	0.61 (0.49)	0.371	0.032
Student	mean (SD)	0.05 (0.22)	0.05 (0.21)	0.603	0.019
Dual student	mean (SD)	0.00 (0.06)	0.01 (0.09)	0.072*	0.065
Parental leave	mean (SD)	0.01 (0.10)	0.01 (0.11)	0.381	0.031
Unemployed, searching	mean (SD)	0.03 (0.17)	0.02 (0.14)	0.097*	0.059
Unemployed, not searching	mean (SD)	0.02 (0.14)	0.02 (0.14)	0.862	0.006
Not gainfully employed	mean (SD)	0.25 (0.44)	0.24 (0.43)	0.350	0.033

Table 7: Balance check across covariates by treatment status

Variable/Category	Unit	Treatment Status		Tests for Difference	
		treatment	control	p-value	SMD
Working hours per week	mean (SD)	23.98 (18.39)	24.55 (18.36)	0.382	0.031
Ethnicity	count (%)			0.478	0.104*
African		5 (0.3)	7 (0.4)		
European/Caucasian		1423 (93.1)	1492 (93.5)		
Mixed		18 (1.2)	24 (1.5)		
Caribbean		2 (0.1)	0 (0.0)		
Latino/Hispanic		6 (0.4)	1 (0.1)		
Middle Eastern		11 (0.7)	11 (0.7)		
East Asian/Southeast Asian		8 (0.5)	5 (0.3)		
South Asia		5 (0.3)	4 (0.3)		
No Answer		26 (1.7)	31 (1.9)		
Other		24 (1.6)	20 (1.3)		
Real estate	mean (SD)	0.43 (0.50)	0.44 (0.50)	0.600	0.019
Religious affiliation	count (%)			0.025**	0.156*
No religion		796 (52.1)	778 (48.8)		
Protestant		263 (17.2)	297 (18.6)		
Roman Catholic		325 (21.3)	362 (22.7)		
Mormon		0 (0.0)	1 (0.1)		
Eastern or Greek Orthodox		23 (1.5)	18 (1.1)		
Jewish		3 (0.2)	9 (0.6)		
Muslim		35 (2.3)	53 (3.3)		
Buddhist		12 (0.8)	2 (0.1)		
Hindu		5 (0.3)	3 (0.2)		
Other		66 (4.3)	72 (4.5)		
Importance of religion	mean (SD)	3.00 (2.13)	3.08 (2.15)	0.286	0.038
Importance of democracy	mean (SD)	6.29 (1.10)	6.24 (1.13)	0.243	0.042
Likelihood voting for					
CDU	mean (SD)	5.01 (3.42)	5.03 (3.37)	0.880	0.005
Grüne	mean (SD)	4.13 (3.40)	3.95 (3.34)	0.134	0.054
SPD	mean (SD)	4.62 (3.20)	4.50 (3.14)	0.295	0.037
Linke	mean (SD)	3.09 (2.85)	3.13 (2.90)	0.734	0.012
SPD	mean (SD)	3.59 (2.85)	3.59 (2.85)	0.972	0.001
AfD	mean (SD)	3.39 (3.52)	3.49 (3.52)	0.428	0.028
Voted in elections	count (%)			0.292	0.056
Yes		1386 (90.7)	1450 (90.9)		
No		129 (8.4)	123 (7.7)		
No answer		13 (0.9)	22 (1.4)		
Political interest	mean (SD)	4.99 (1.59)	4.93 (1.61)	0.318	0.036
Discussing politics with others	mean (SD)	2.95 (1.07)	3.01 (1.11)	0.147	0.052
Consumption of political news	count (%)			0.054*	0.118*
Every day		497 (32.5)	555 (34.8)		
Often		246 (16.1)	234 (14.7)		
Regularly		382 (25.0)	345 (21.6)		
Sometimes		319 (20.9)	382 (23.9)		
Never		74 (4.8)	65 (4.1)		
I am not sure		10 (0.7)	14 (0.9)		
How democratic is Germany	mean (SD)	59.99 (25.25)	58.77 (25.28)	0.179	0.048
Partisanship	count (%)			0.628	0.035

Table 7: Balance check across covariates by treatment status

Variable/Category	Unit	Treatment Status		Tests for Difference	
		treatment	control	p-value	SMD
No strong supporter of any party		721 (47.2)	780 (48.9)		
Strongly supports one party		646 (42.3)	654 (41.0)		
Strongly supports two or more parties		161 (10.5)	161 (10.1)		

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

P-values come from two-sample t-tests for continuous variables and chi-square tests for categorical variables.

Standardized Mean Differences (SMD) measure effect size:

Good balance: $SMD < 0.1$:

Small difference (minor imbalance): SMD from 0.1 to 0.2 *

Moderate or large imbalance: $SMD > 0.2$ **