Science under Threat? A Natural Experiment in Economics

Dominic Rohner, Oliver Vanden Eynde and Philine Widmer*

February 2026

Abstract

Academic freedom has come under growing strain across the world. To study whether and how academics react to political pressure, we exploit a natural experiment: the U.S. government's "blacklist" of undesirable words released in early 2025. We find that the release of this list leads to a sharp reduction in the use of banned words in sensitive contexts among economists working at universities that rely heavily on NSF funding. The drop is particularly marked for content related to gender, race, and environment. Our findings are consistent with scholars responding strongly to political pressure through career incentives.

JEL Classification: D73, I23, O38.

Keywords: Censorship, Science, Academic Freedom, Science Funding.

*Rohner: Geneva Graduate Institute, University of Lausanne and CEPR (email: dominic.rohner@graduateinstitute.ch); Vanden Eynde: Paris School of Economics, CNRS, and CEPR (email: o.vandeneynde@psemail.eu); Widmer: Paris School of Economics (email: philine.widmer@gmail.com). This research was supported by EUR grant ANR-17-EURE-0001 at PSE. Violaine D'Ortona provided excellent research assistance.

1 Introduction

In recent years, academic freedom has been increasingly challenged, in many countries and by actors of various political orientations. The 2025 Academic Freedom Index (Kinzelbach et al., 2025) shows that in the 10-year period between 2014 and 2024, academic freedoms saliently decreased in 34 countries (including heavyweights like the United States, India, and Russia), while only increasing in a mere 8 countries (all of which are relatively less populous).

Whether pressure on academic freedom really translates into biases in the choice of research content crucially depends on how scholars react. Do academics yield to political pressures or are their choices of research topics guided by scientific interest alone, even if this means potentially "paying a price" in terms of career advancement? Václav Havel, former Czechoslovak dissident and ex-president of Czech Republic illustrates this dilemma in his 1978 essay "The Power of the Powerless" (Havel, 2009), where he tells the story of a greengrocer displaying the slogan "Workers of the world, unite!" in his shop window for the sole purpose of avoiding trouble, while accepting to "live within a lie". The stakes are high, as populist leaders can hollow out democratic institutions if enough people adopt the greengrocer's attitude. In principle, academia could provide a shielded ivory tower environment. Yet, in reality, it has come under mounting political pressure, as illustrated by the measured decline in academic freedom. Today, rock-solid scientific evidence is increasingly challenged by political dogma on topics such as climate change and vaccines.

Studying scientific (self-)censorship is methodologically challenging, as there is rarely exogenous variation in academic freedom that can be easily measured. Furthermore, long publication delays make it difficult to study the timing of changes in academic freedoms, and it is notoriously arduous to disentangle scholars' behavior from potential biases in the publishing process. However, we can draw on a unique natural experiment that allows us to address all of the above methodological roadblocks.

In particular, we investigate how the U.S. government's "black list" of undesired words has affected the choice of research topics in economics. In February 2025, U.S. federal agencies like the National Science Foundation (NSF) and the Centers for Disease Control and Prevention (CDC) began enforcing lists of "forbidden words"—including terms such as "diversity," "equity," "inclusion," "transgender," and "climate change"—in research documents posted by these agencies and grant applications (Yourish et al., 2025). Researchers using these words came under scrutiny and saw their existing research grants or future grant applications jeopardized. By January 2026, 1,996 NSF grants have been canceled or suspended (Kozlov et al., 2026). Our analysis exploits the sharp discontinuity of U.S. government policy and the heterogeneity between scholars more or less exposed to U.S. governmental pressure, based on the reliance of their institutions on NSF funding. We can draw on the particularly vibrant working paper culture in economics: unlike in most disciplines, it is standard practice to circulate working papers before they are peer-reviewed. This means we can precisely identify the immediate impact of the U.S. policy change. It also allows us to focus solely on scholar behavior, while filtering out potential impacts from the publication process or the review process for new grants. Beyond providing a well-suited case study, economics is also fertile ground for our study because recent work has found that this discipline has a substantial impact on policymakers' decisions (Hjort et al., 2021).

Our results show that the release of the list of banned words leads to a sharp reduction in the use of these words in a sensitive context by economists working in universities that rely heavily on NSF funding, compared to those working in institutions that are less exposed to federal funding. In particular, we observe a 7.5 percentage point drop in the probability of using any of the banned words in the context of gender, race, or environment. We also observe this drop when we filter out any time-invariant author characteristics through author fixed effects. The marked decline in the use of banned words in research papers from NSF-intensive US institutions holds across papers from a range of reference groups: low NSF-intensive US institutions, Research Council-intensive UK universities, and a reference category consisting of all other institutions. The reduction in use is particularly marked for content related to gender, race, and environment. Importantly, the reduction in word use is not driven by research papers that acknowledge NSF funding, which rules out that our findings reflect a direct effect. Indeed, the U.S. government has no direct control over the content of working papers (even if research is NSF-funded). The ban is directly relevant only to existing and future NSF grants. In this context, the sudden impact of the word ban, including for research that is not NSF-funded, can only be explained by self-censorship. The response we document may reflect a strong degree of anticipation about how one's current research might affect one's career paths and funding opportunities amid severe political pressure on academic research funding.

The current work relates to several relevant strands of the literature. First, the pursuit of economic research is a topic of research in its own right (e.g., Hamermesh, 2013; Fourcade et al., 2015; Garg and Fetzer, 2025a). A series of studies has investigated publication, citation, and implementation bias. The former, publication bias, refers to the fact that – even if access to data is secured and the researcher has designed a credible empirical strategy – significant results are more likely to be published, which may induce "p-hacking" by scholars (Brodeur et al., 2016, 2020; Kasy, 2021).[1] Underpowered research designs could further skew evidence towards extreme findings (Ioannidis et al., 2017). Strategic behavior not only shapes which scientific findings are published but also influences citation practices. Rubin and Rubin (2021) show the importance of such citation bias, as research from the Journal of Business saw a sharp fall in citations relative to other research following the discontinuation of the journal. Another well-known type of bias is the so-called implementation bias, which refers to the fact that randomized control trials (RCTs) could be subject to external validity biases (Chassang et al., 2012; Banerjee et al., 2017).

This existing work on scientific publications focuses on a set of biases related to the publication process, but it is not directly concerned with political pandering or reactions to political pressure. There is more work on political censorship and propaganda in the media (e.g., Yanagizawa-Drott, 2014; Guriev and Treisman, 2019; Widmer, 2024). However, the

[1] Retractions of published articles are relatively rare, but they can play an important role in correcting scientific knowledge when research findings are later identified as wrong (Alabrese, 2022).

institutional structure and societal role of scientific research differ from those of the press, which has a more natural proximity to politics. Still, experiences of scientific censorship appear to be relatively common in academia. A survey among US faculty in 2022 found that 4% had been disciplined or threatened with discipline because of their research, academic talks, or non-academic publications (Honeycutt et al., 2023).[2] Clark et al. (2023) argue that such scientific censorship is often driven by other researchers, who may be motivated by self-protection, benevolence toward peers, and prosocial concerns for the well-being of social groups. Biases in science funding, which may be at the root of self-censorship, are also understudied. One recent exception is by Furnas et al. (2026), who show that funders penalize US-Chinese scientific collaborations compared to US-German teams. More broadly, Iaria et al. (2018) demonstrate that the collapse of international scientific cooperation during and after World War I reduced the output of researchers cut off from foreign frontier knowledge, illustrating how political disruptions can reshape the production of science.[3] Some studies also measure researchers' political biases. Garg and Fetzer (2025b) find that US academics who express themselves on Twitter diverge from general public opinion in topic focus, while Alabrese et al. (2024) show that political expression by scholars on social media may undermine their credibility. This recent work does not speak to the extent to which scientific content itself is biased. Addressing that question, Jelveh et al. (2024) and Borjas and Breznau (2026) find that economists are influenced in their research by their individual political attitudes. While there is evidence that institutional and political forces can change the behavior of institutional actors – for example, Ash et al. (2025) show that exposure to economics training causally changed judges' language and rulings, while Grosjean et al. (2023) show that Trump rallies increased racial bias in policing[4] – to the best of our knowledge, there is no work on how exogenous political shocks affect scholars' choices of research

[2]In addition, researchers may adjust the way they communicate about science to the tastes of their audience (Ratcliff et al., 2023).

[3]Waldinger (2012) also exploits political interference in academia – the dismissal of scientists in Nazi Germany – to study peer effects in science.

[4]In related work, Vanden Eynde et al. (2018) find that street-level policemen in Kenya misbehave more when their ethnic group holds political power.

content, potentially biasing the production of scientific knowledge. The novelty and value added of our current paper is precisely to address this gap in the literature.

The remainder of this paper is structured as follows. Section 2 presents the data and methods, whereas Section 3 depicts the findings. Mechanisms and channels of transmission are studied in Section 4. Finally, Section 5 concludes. Further data description and additional empirical results are relegated to the Appendix.

2 Data and Methods

In what follows, we shall describe the data used and methods applied.

2.1 Data

Working papers. We collect the universe of working papers published by the National Bureau of Economic Research (NBER) and the Centre for Economic Policy Research (CEPR) between January 2020 and December 2025.[5] NBER papers are obtained via the NBER API, which provides structured metadata including title, authors (with profile URLs), abstract, publication date, JEL codes, and program affiliations.[6] CEPR papers are collected by scraping the CEPR discussion paper listing pages, extracting titles, authors, abstracts, dates, keywords, and JEL codes. For both sources, we download the full PDF of each paper and extract the complete text using `pdfplumber`.

Since some papers are published in both series, we de-duplicate by matching on lowercased titles. When a paper appears in both NBER and CEPR, we retain the CEPR version. After de-duplication, our dataset contains 14,412 unique papers: 8,007 from NBER and 6,405 from CEPR (Table A2).[7]

[5]Garg and Fetzer (2025a) also rely on a sample of CEPR and NBER working papers. They study changes in economic research methodologies over time.

[6]The API endpoint is `https://www.nber.org/api/v1/working_page_listing/`.

[7]It is important to note that both NBER and CEPR are independent, non-partisan organizations run by academics for academics. While formally working papers are approved by program directors, *de facto* the content of working papers is typically solely up to the affiliated researcher.

Author affiliations and funding data. We extract author affiliations from the second page of each PDF using Claude Sonnet, which identifies each author's university affiliation(s) while stripping department names, addresses, and research bureau affiliations (e.g., "and NBER"). When an author lists multiple universities, the matching step maps the combined string to the institution found in the funding reference list; if both appear, the first match is retained.[8] For authors with missing affiliations, we fill in the institution using the same author's affiliation from another paper.

We match US universities to the Higher Education Research and Development (HERD) Survey (FY2023), which reports university-level R&D expenditure by funding source. For UK universities, we use the Higher Education Statistics Agency (HESA) Table 5 (academic year 2022–23), which provides analogous data on Research Council funding. University matching is performed using Claude Sonnet, which maps author-reported institution names to the canonical names in the HERD list (344 US institutions) and the HESA list (224 UK institutions).

Our key treatment variable, *High-NSF*, indicates universities whose NSF funding share of total R&D expenditure exceeds the median across all US universities in the HERD data. At the paper level, a paper is classified as High-NSF if the majority of its US-affiliated authors are at High-NSF universities; at the author level, the classification is based on the individual author's university. For the UK, we construct an analogous indicator based on the Research Council (RC) funding share from the HESA data; at the paper level, a paper is classified as High-RC if the majority of its UK-affiliated authors are at above-median-RC-share universities. Of the 14,412 papers, 10,030 (69.6%) have at least one US-affiliated author matched to HERD funding data, and 2,156 (15.0%) have at least one UK-affiliated author matched to HESA data. Paper-level summary statistics are in Table A2; author-level summary statistics are in Appendix Table A3.

[8] Affiliations are extracted independently from each paper's cover page, so authors who change institutions during the sample period are assigned different affiliations for different papers.

Targeted terms. We construct a list of terms flagged by U.S. federal agencies following the executive orders of February 2025, as documented by Yourish et al. (2025). Starting from the 174 terms documented by Yourish et al. (2025), we add morphological variants (e.g., "barrier" and "barriers," "bias" and "biased") to arrive at 197 base terms including "diversity," "equity," "inclusion," "transgender," "climate change," "gender identity," and "racial justice" (see Appendix Table A1 for the complete list). We further expand the list to 224 search patterns by adding UK spellings (e.g., "marginalise"), hyphenation variants (e.g., "anti-racism" and "antiracism"), and compound forms (e.g., "breastfeeding people").

Several expressions in the list are nested – for instance, "gender," "gender-based," and "gender-based violence" are all separate entries. To avoid double-counting, we sort all terms by length in descending order before constructing a single regular expression. Python's `re.findall()` function tries alternatives left to right and consumes matched text, so the longest matching expression is always counted first: "gender-based violence" is recorded as one match, not three. This procedure is applied identically to both abstract-level and full-text counts. On average, papers contain 1.1 targeted terms in the abstract and 82.5 in the full text (Table A2).

Context classification. Many targeted terms are ambiguous in economics – "equity" may refer to financial equity, "bias" to statistical bias, and "race" to a competitive contest. To distinguish substantive from incidental uses, we classify each occurrence of a targeted term in the abstract into one of five thematic categories using Claude Haiku. The classification prompt is detailed in Appendix A.1.

Of the 5,477 papers with at least one targeted term in the abstract, 19.8% contain at least one term used in a Gender context, 9.4% in a Race context, 9.3% in an Environment context, 66.1% in an Economic inequality context, and 8.9% in a Cannot say or other context. Since each word is classified separately, papers with multiple targeted terms can appear in more than one category. Figure A1 displays odds-ratio word clouds for the four categories most relevant to our analysis.[9] The Gender cloud is dominated by terms like "parental leave," "child penalty," "gender pay," and "gender norms"; Race by "racial bias," "black children," "white households," and "racial disparities"; Environment by "carbon," "emission," "carbon tax," and "clean energy"; and Economic inequality by "capital income," "skill premium," "wealth inequality," and "inequality dynamics."

Our main analysis focuses on papers classified as Gender, Race, or Environment (GRE) – the categories most closely associated with what the Trump administration characterizes as "woke" and which directly map to the stated rationale behind the word ban (Yourish et al., 2025). We show robustness of all results using simple abstract word counts and full-text word counts, which yield qualitatively similar but noisier estimates. Figure 4 examines heterogeneity across individual content categories. As a further robustness check, we classify targeted terms in the full paper text – not just abstracts – to verify that the abstract-level results reflect substantive content changes rather than superficial word substitution. Because classifying every targeted term across entire papers is prohibitively costly with the Anthropic API, we use the open-source Qwen3-32B model (accessed via the Groq API) for this full-text classification.[10] Appendix Figures A5 and A9 confirm that results are robust to this alternative classification model and text source.

NSF and SES mentions. To test whether our effects are driven by authors who received NSF funding themselves versus those who merely work at NSF-dependent institutions, we flag papers that mention "NSF" or "SES" (the Division of Social and Economic Sciences, the NSF division most relevant to economics) in their acknowledgment sections. We use these flags to conduct analyses that exclude directly funded papers (Figure 5).

[9] Each word is scaled by its log-odds ratio of appearing in an abstract classified in a given category relative to abstracts not in that category, computed from unigram and bigram count vectorizations of all classified abstracts.

[10] All classifications – whether by Claude or Qwen – are based on publicly available abstracts or short text excerpts (∼200 characters) surrounding each targeted term, not on full paper texts.

2.2 Methods

The identification strategy of the current paper is based on exploiting heterogeneous exposure to an (unanticipated) policy shock, unfolding in January/February 2025. In what follows, we first shall demonstrate that the "treatment group" of highly affected scholars (i.e. those from highly NSF-funding dependent U.S. universities) had a comparable pre-trend with the "control group" of less affected scholars. Then, as the "black list" of undesired words has been issued in February 2025, one may expect a different reaction from the treatment versus the control group, which is what we investigate empirically.

Note that we focus on the *relative* reactions of more versus less exposed scholars with respect to a sharp policy change. Our setting allows us to detect to what extent academics react to political pressure, but we are not able to make general statements about the absolute levels of self-censorship, i.e., our data would not allow us to know if, during the pre-treatment period, scholars have self-censored themselves, and if yes, in what direction. Importantly, however, given that our identification strategy is based on heterogeneous effects of sharp changes, we do not need such information on absolute levels. Indeed, in line with our research question, our setting allows us to measure the extent to which academics alter their research content in response to political pressure.

Event study. Our main specification is a half-yearly event study estimated as a linear probability model:

$$Y_{it} = \sum_{\substack{s=2020\text{H1} \\ s\neq 2024\text{H2}}}^{2025\text{H2}} \beta_s \cdot \mathbf{1}[t=s] \times \text{HighNSF}_i + \gamma_t + \delta \cdot \text{Source}_i + \varepsilon_{it} \tag{2.1}$$

where Y_{it} is a binary indicator equal to one if paper i published in half-year t contains any targeted term in a Gender, Race, or Environment context in its abstract, HighNSF_i indicates whether the majority of the paper's US-affiliated authors are at above-median-NSF-share universities, γ_t are half-year fixed effects, and Source_i is a NBER/CEPR indicator. The reference period is 2024H2, the last half-year before the policy change. The coefficients of interest, β_s, trace the differential evolution of censored-term usage between High-NSF and Low-NSF papers over time.

At the paper level, we use heteroskedasticity-robust standard errors. At the author level, the unit of observation becomes the author-paper pair: a paper with K co-authors contributes K observations. We estimate

$$Y_{ijt} = \sum_{\substack{s=2020\text{H1} \\ s\neq 2024\text{H2}}}^{2025\text{H2}} \beta_s \cdot \mathbf{1}[t = s] \times \text{HighNSF}_j + \alpha_j + \gamma_t + \delta \cdot \text{Source}_i + \varepsilon_{ijt} \tag{2.2}$$

where j indexes authors, i papers, and t half-years. HighNSF_j is defined at each author's own university, α_j are author fixed effects, and standard errors are clustered by university. We also estimate pooled pre- and post-treatment averages and report the difference between them. As robustness checks, we estimate specifications with count dependent variables (full-text word counts) using Poisson pseudo-maximum likelihood (PPML).

Reference group comparisons. To assess robustness across different control groups, we estimate pooled difference-in-differences specifications:

$$Y_{it} = \beta \cdot \text{Post}_t \times \text{Treated}_i + \delta_1 \cdot \text{Post}_t + \delta_2 \cdot \text{Treated}_i + \delta_3 \cdot \text{Source}_i + \varepsilon_{it} \tag{2.3}$$

where $\text{Post}_t = \mathbf{1}[\text{year} \geq 2025]$ and Treated_i is a binary indicator that varies across comparisons: US High-NSF vs. US Low-NSF, US High-NSF vs. UK High-RC, US High-NSF vs. all UK, US High-NSF vs. all other, US (all) vs. UK (all), and two placebo comparisons (US Low-NSF vs. UK Low-RC and UK High-RC vs. UK Low-RC). The same logic extends to all pairwise comparisons shown in Figure 3. Paper-level regressions use robust standard errors; author-level regressions replace i with ij subscripts and add author fixed effects α_j as in equation (2.2), with standard errors clustered by university.

Heterogeneity measures. To study whether the self-censorship response varies across author subgroups, we construct three author-level characteristics. *Seniority* is measured as years since PhD, computed from PhD graduation years collected via web searches for each author (Appendix A.4). The cutoff at 12 years corresponds to the author-paper-level median; seniority is time-varying, recomputed for each paper's publication year, and the sample is restricted to authors who are active in the pre-treatment period. *Ethnicity* is predicted from the US Census 2010 surname data file, which contains approximately 162,000 surnames with associated race/ethnicity probabilities. Each author's last name is looked up in the Census file, and the highest-probability category is assigned as the predicted ethnicity. We group predictions into three categories: White, Minority (comprising Asian/Pacific Islander, Hispanic, Black, and American Indian), and Surname not in Census (predominantly non-English names). The Census match rate among US-affiliated authors is approximately 77% (Appendix A.3). *Gender* is classified as male or female from authors' first names using Claude Haiku (Appendix A.2).

3 Results

In what follows, we first discuss key features of the raw data, then present the main econometric results. In particular, Figure 1 provides a first look at the data. The left panel displays the most frequent targeted terms found in paper abstracts. The word cloud is dominated by broadly used terms such as "inequality," "gender," "bias," "equity," and "women," alongside more specific expressions like "climate change" and "institutional." The right panel shows that 1,991 of the 14,412 papers in our sample (13.8%) contain at least one targeted term used in a Gender, Race, or Environment context – the categories most closely associated with the stated rationale behind the word ban. We now examine how the prevalence of this sensitive content evolves differentially by institutions' exposure to NSF funding.

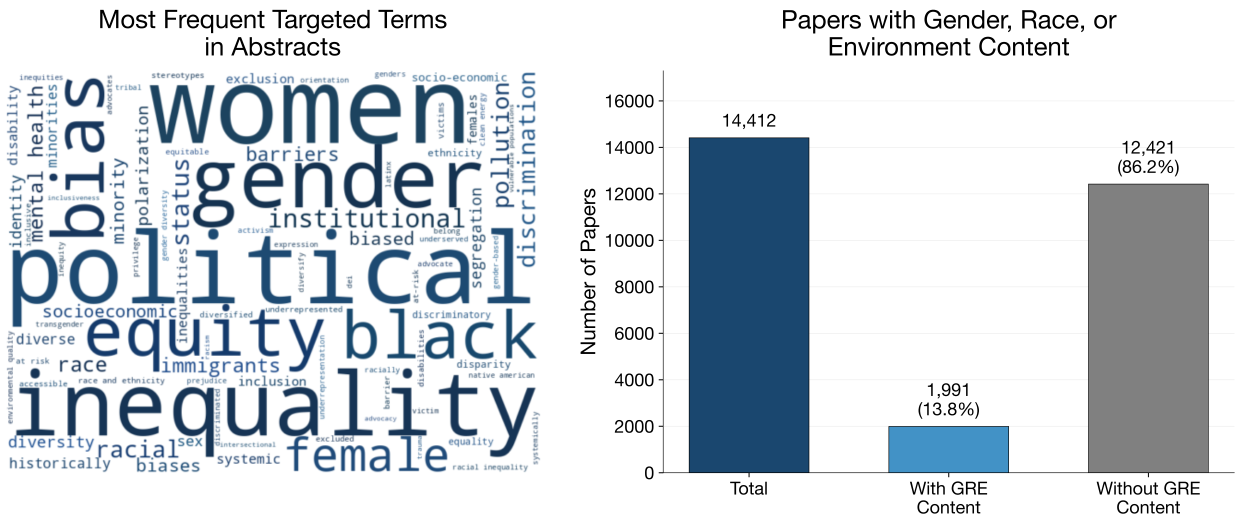


Figure 1: Targeted Terms and Gender, Race, and Environment Content in Economics Working Papers

Notes: The left panel displays the most frequent targeted terms found in abstracts of NBER and CEPR working papers (January 2020–December 2025), with word size proportional to frequency. The right panel shows the number of papers containing targeted terms in a Gender, Race, or Environment context, as classified by Claude Haiku based on the surrounding text in the abstract. Of the 14,412 papers, 1,991 (13.8%) contain at least one targeted term in one of these three contexts.

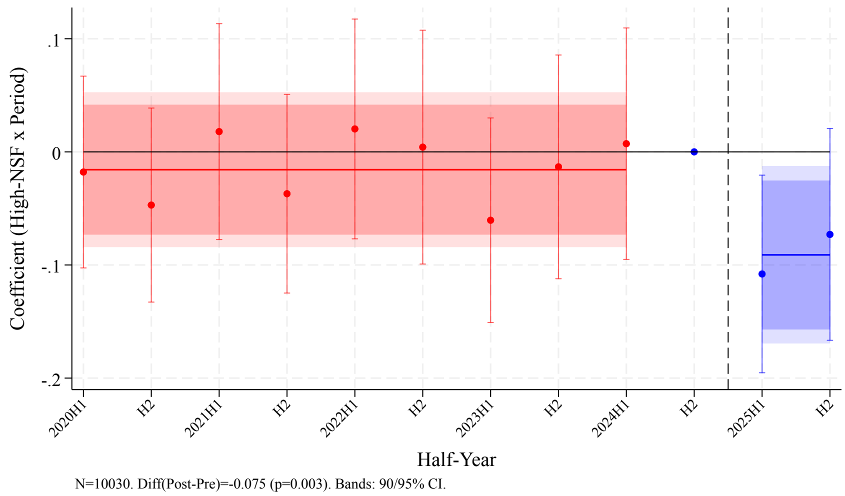
Event study. Figure 2 plots half-yearly event study coefficients from equation (2.1), comparing the evolution of Gender, Race, and Environment content prevalence between US High-NSF and Low-NSF institutions. The pre-treatment coefficients fluctuate around zero with no discernible trend, supporting the parallel-trend assumption. After the policy change, the 2025H1 coefficient at the paper level is -0.108 ($p = 0.015$), and the pooled difference between post- and pre-treatment averages is -0.075 ($p = 0.003$). This corresponds to a 7.5 percentage point drop in the probability of using any of the banned words in a gender, race, or environment context. At the author level, where author fixed effects absorb time-invariant individual characteristics, the pooled post-pre difference is -0.046 ($p = 0.059$), suggesting that the decrease in banned words is driven by changes in individual researchers' behavior rather than compositional shifts in who publishes. Again, the effect is sizable: it corresponds to a 4.6 percentage point drop in the probability of using a banned word in a sensitive context. Appendix Figure A2 plots the raw share of GRE-related abstracts by half-year for High-NSF and Low-NSF institutions. The two groups follow parallel trends in the pre-period, with High-NSF institutions at a consistently higher level; after the policy change, the High-NSF share drops while Low-NSF institutions' share remains largely stable.

Reference group comparisons. Figure 3 presents pooled difference-in-differences estimates from equation (2.3) across several treatment-control comparisons. At the paper level, the coefficient is -0.077 ($p = 0.002$) for US High-NSF versus Low-NSF, -0.106 ($p = 0.099$) versus UK High-RC, -0.074 ($p = 0.043$) versus all UK, and -0.069 ($p = 0.004$) versus all other institutions. At the author level with author fixed effects, the primary comparison yields -0.046 ($p = 0.064$); the other treatment comparisons are sign-consistent but less precisely estimated. Both placebo comparisons – US Low-NSF versus UK Low-RC, and UK High-RC versus UK Low-RC – are insignificant at both levels, consistent with the identification strategy.

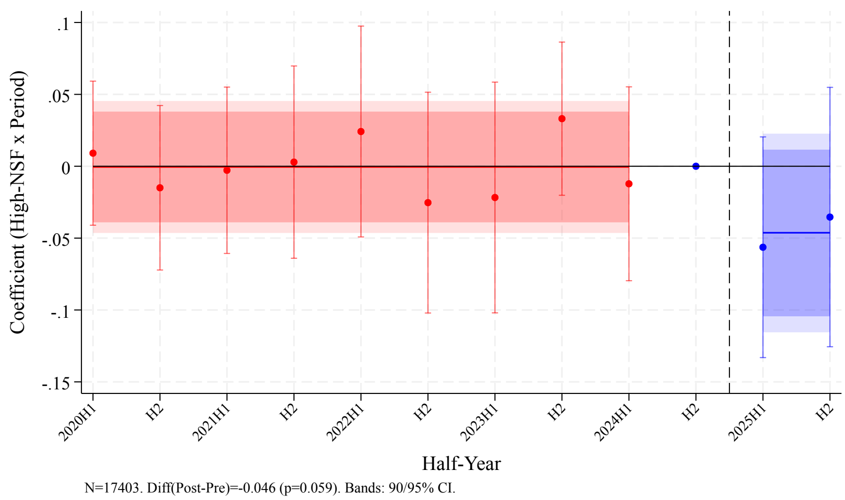
Robustness. We probe the robustness of both sets of results – event studies and reference group comparisons – along several dimensions. While the first set of robustness figures will focus on event studies (Appendix Figures A3–A6), the second series of robustness figures will focus on coefficient plots for reference group comparisons and A7–A12).

Intensive margin. First, we move from the extensive margin to the intensive margin by counting the number of GRE-classified words in each abstract rather than using a binary indicator (see Appendix Figures A3 and A7). PPML estimates confirm a significant decline in GRE word counts at both the paper level (-0.565, $p = 0.014$) and the author level (-0.473, $p = 0.001$), and reference group comparisons are similarly robust.

Without context classification. Second, we drop the context classification step entirely and simply flag whether any targeted term appears anywhere in the abstract (see Appendix Figures A4 and A8). Estimates remain qualitatively similar but are attenuated, consistent with many targeted terms having technical uses in economics. So, the classification step sharpens measurement by isolating politically sensitive uses.



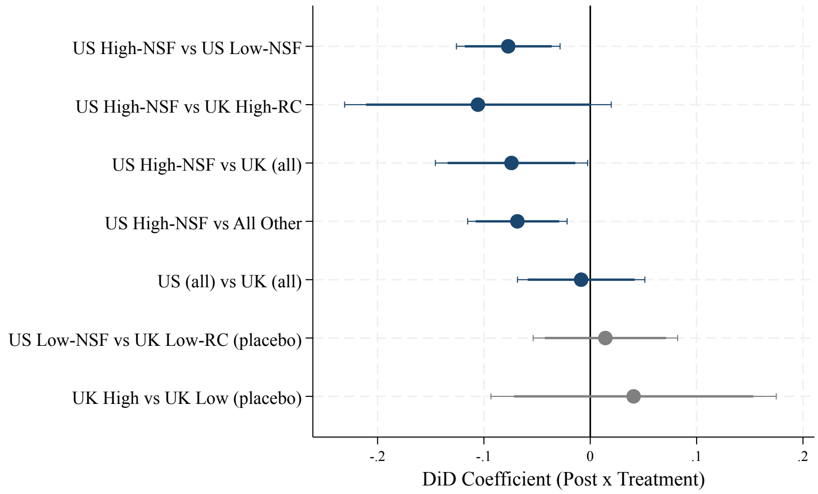
(a) Paper-Level: Gender+Race+Environment Content (Reference: 2024H2)



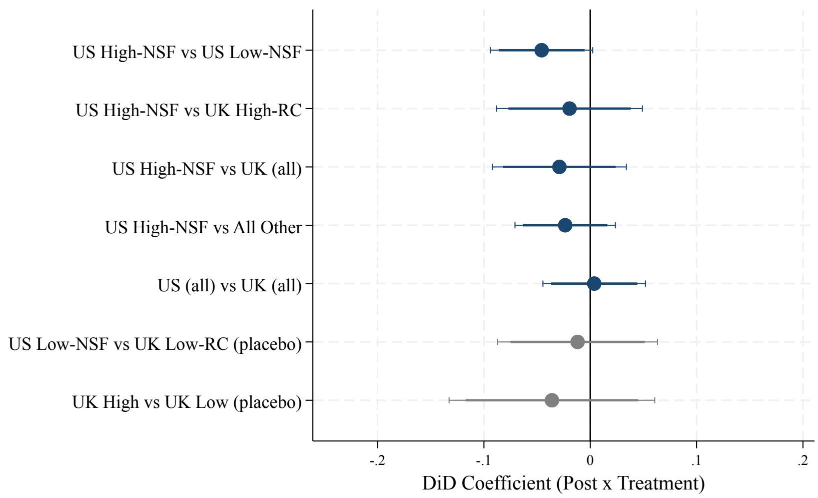
(b) Author-Level: Gender+Race+Environment Content (Reference: 2024H2)

Figure 2: Half-Yearly Event Study: Effect on Gender+Race+Environment Content

Notes: Figures plot half-yearly difference-in-differences coefficients comparing US High-NSF institutions to US Low-NSF institutions, with 2024H2 as the reference period. High-NSF universities are those whose NSF funding share of total R&D expenditure (HERD survey, all fields) exceeds the cross-university median. Low-NSF universities are those at or below the median. At the paper level, a paper is classified as High-NSF if the majority of its US-affiliated authors are at High-NSF universities. The dependent variable is binary, indicating whether the paper contains any censored word in a Gender, Race, or Environment context. Context is assigned by Claude Haiku based on the surrounding text in the abstract. Red shading indicates pre-treatment periods; blue shading indicates post-treatment periods. Darker bands represent 90% confidence intervals; lighter bands represent 95% confidence intervals. Horizontal lines show pooled pre- and post-treatment averages. The upper panel shows paper-level estimates with source fixed effects (robust standard errors). The lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). Linear probability model estimates. See Appendix Figures A3–A6 for robustness checks varying the outcome measure (GRE word counts, any censored word in abstracts, Qwen3 full-text GRE word counts, and full-text censored word counts).



(a) Paper-Level: Gender+Race+Environment Content



(b) Author-Level: Gender+Race+Environment Content

Figure 3: Treatment Effects Across Reference Groups: Gender+Race+Environment Content

Notes: Figures display pooled difference-in-differences coefficients for various treatment-control comparisons. The dependent variable is binary, indicating whether a paper contains any censored word in a Gender, Race, or Environment context (classified by Claude Haiku based on surrounding text). High-NSF universities are those whose NSF funding share of total R&D expenditure (HERD survey, all fields) exceeds the cross-university median. Low-NSF universities are at or below the median. At the paper level, a paper is classified as High-NSF if the majority of its US-affiliated authors are at High-NSF universities. The top five rows (navy) compare US High-NSF authors against progressively broader control groups – US Low-NSF, UK High-RC, all UK, and all non-High-NSF authors – as well as all US versus all UK; the bottom two rows (gray) are placebo tests comparing groups that should be unaffected by NSF policy. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Upper panel: paper-level estimates with robust standard errors; lower panel: author-level with author fixed effects and standard errors clustered by university. Linear probability model estimates. See Appendix Figures A7–A10 for robustness checks varying the outcome measure; Appendix Figure A11 for a 2024-only pre-treatment window; and Appendix Figure A12 for robustness to controlling for Red/Blue state × Post. Tabular versions are in Appendix Tables A4 and A5.

Full paper. Third, we move from abstracts to full paper texts using both a GRE classification based on Qwen3-32B (Appendix Figures A5 and A9) and raw counts of targeted words (see Appendix Figures A7 and A10).[11] Full-text measures are inherently noisier: with an average of 82.5 targeted terms per paper compared to 1.1 in abstracts, marginal changes are harder to detect and less likely to reflect deliberate content shifts. Nevertheless, estimates remain negative, though less precisely estimated.

Time window. As a separate identification check, restricting the pre-treatment window to 2024 alone produces sign-consistent estimates, even if some reference group comparisons become imprecise (Appendix Figure A11).[12]

Additional controls. Finally, adding an interaction between an indicator for universities in states that voted Republican in 2024 and the post-treatment dummy leaves all estimates essentially unchanged (Appendix Figure A12), ruling out that the NSF exposure effect is driven by the state-level political climate.

4 Mechanisms

The preceding results establish that researchers at NSF-dependent institutions reduced their use of politically sensitive terms after the policy change. We now explore four dimensions of this response. First, we decompose the aggregate Gender, Race, and Environment effect by content category to assess whether the decline is broad-based or driven by a single topic. Second, we test whether the effect operates through direct financial incentives – by excluding papers that acknowledge NSF funding – or reflects a broader response. Third, we examine whether the self-censorship response varies across author subgroups defined by seniority, predicted ethnicity, and predicted gender. Fourth, we investigate the impact on academic output (paper length and number of working papers).

[11] All full-text specifications control for the number of pages.

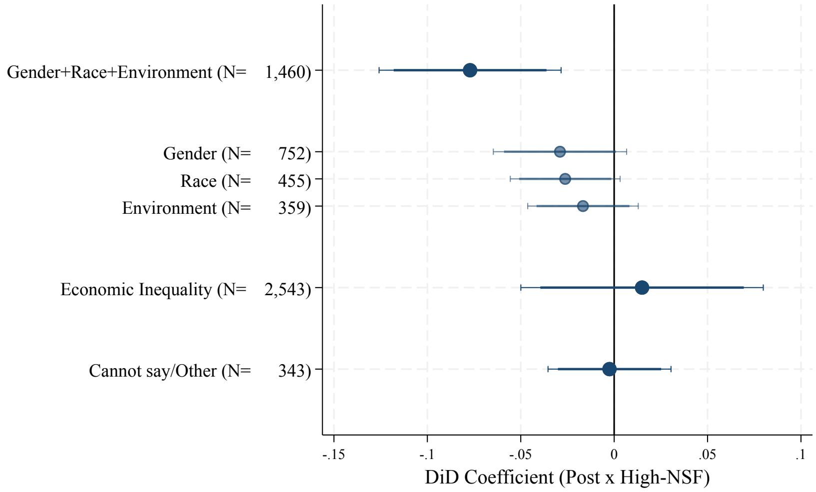
[12] In a sample restricted to 2024–2025, author fixed effects require authors who publish in both years, leaving too few observations for reliable inference. This is why we only present paper-level results here.

Content heterogeneity. Figure 4 disaggregates the treatment effect by context category. The combined Gender, Race, and Environment effect is not dominated by any single subcategory: all three display negative coefficients of comparable magnitude at both the paper and author levels. The lower statistical significance of the individual categories relative to the aggregate is expected, as each contains mechanically fewer observations. By contrast, the Economic Inequality and Cannot-say/Other categories show near-zero and insignificant effects, indicating that the decline is concentrated in the politically sensitive content targeted by the word ban rather than reflecting a general reduction in research output.

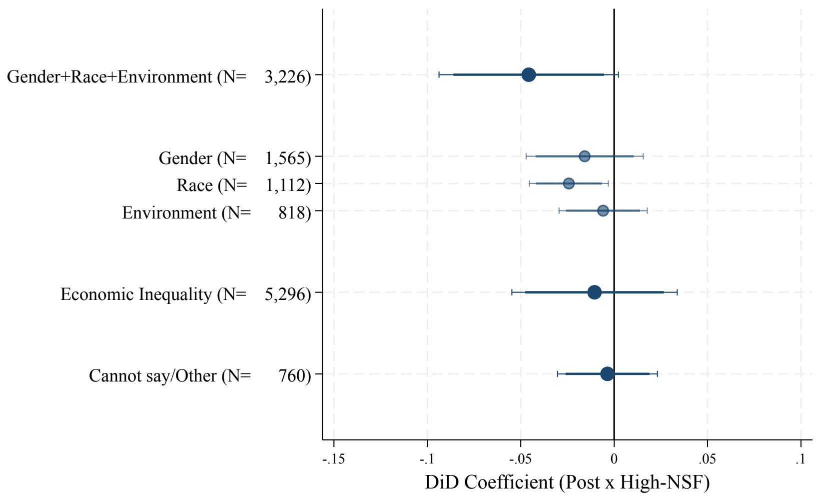
Direct effects of NSF funding. Our treatment variable captures university-level dependence on NSF funding rather than whether a specific paper or author received an NSF grant. Authors at High-NSF institutions face a more uncertain funding environment regardless of their personal grant portfolio. Figure 5 shows that the treatment effect persists when excluding papers that mention NSF or SES funding in their acknowledgment sections.[13] At the paper level, the coefficient moves from -0.071 in the full sample to -0.073 after exclusion; at the author level, from -0.048 to -0.038, remaining directionally consistent. Since the word ban applies directly only to federal grant documents and agency communications – not to working papers – the reduction in the use of targeted terms in research that does not acknowledge NSF funding points to wide-ranging self-censorship. Researchers at NSF-dependent institutions appear to anticipate that their current research output may affect future funding and career prospects, and adjust their research content accordingly.

Heterogeneity by authors. We investigate whether the self-censorship response varies across author subgroups defined by seniority, predicted ethnicity, and gender. Figure 6 reports the DiD coefficient (Post $\times$ High-NSF) estimated separately for each subsample of US authors, alongside pre-treatment level differences (in grey). We find no statistically

[13] We flag both terms because authors differ in how they refer to their grants, and the Division of Social and Economic Sciences (SES) is the NSF division most relevant to economics.



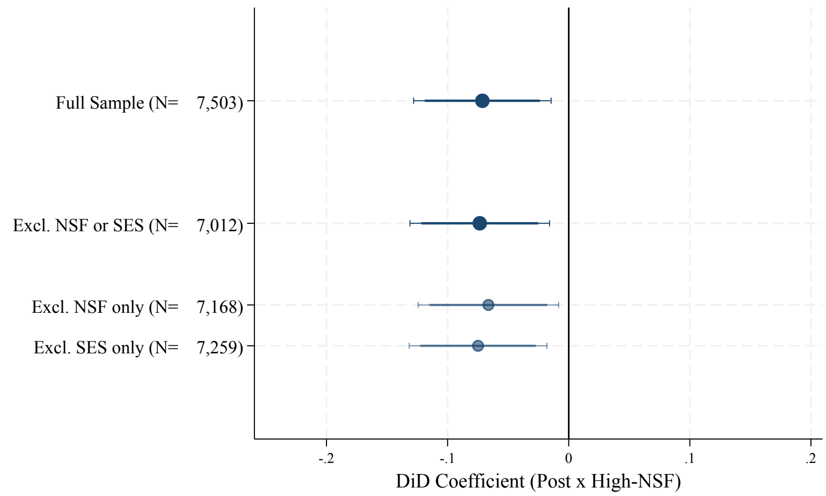
(a) Paper-Level: US High-NSF vs Low-NSF



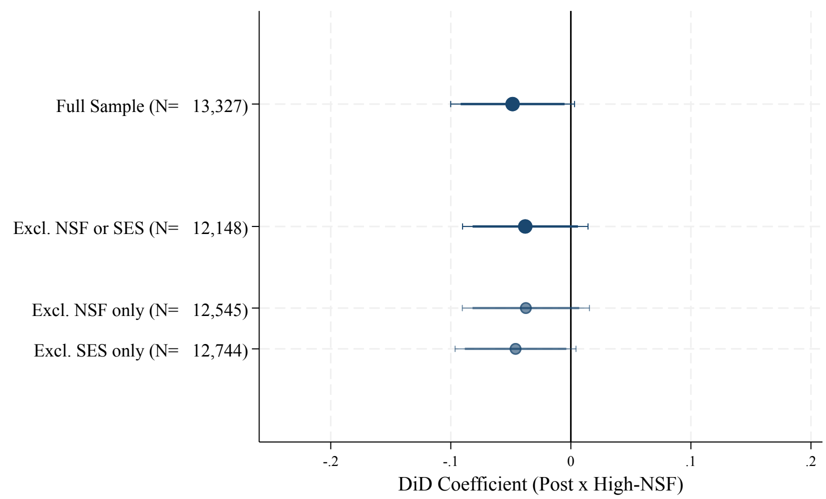
(b) Author-Level: US High-NSF vs Low-NSF

Figure 4: Treatment Effects by Context Category

Notes: Figures display difference-in-differences coefficients separately for papers where censored words appear in different contextual categories. Context is assigned by Claude Haiku based on the surrounding text in the abstract. Papers can appear in multiple categories if they contain words in different contexts. The top row shows the combined "Gender+Race+Environment" category, with its three subcomponents shown directly below (in lighter markers, indented). These three categories represent the most socially salient content areas and provide a robustness check against technical uses of censored terms (e.g., "equity" used in financial contexts). "Economic Inequality" and "Cannot say/Other" follow below. Upper panel shows paper-level estimates with robust standard errors; lower panel shows author-level estimates with author fixed effects. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Linear probability model estimates. Standard errors clustered by university (author level) or robust (paper level). Tabular versions are in Appendix Tables A15 and A16.



(a) Paper-Level: Gender+Race+Environment Content



(b) Author-Level: Gender+Race+Environment Content

Figure 5: Results for NSF/SES Funding Exclusion: Gender+Race+Environment Content

Notes: Figures display difference-in-differences coefficients (US High-NSF vs US Low-NSF) for the full sample and for subsamples excluding papers that mention NSF or SES funding in the acknowledgment sections. The top row shows the baseline effect on the full sample. The second row excludes papers mentioning either "NSF" or "SES." The two rows below (in lighter markers) show the subcomponents: excluding only papers mentioning "NSF" and excluding only papers mentioning "SES." Papers are excluded if they mention either term because different authors refer to their NSF grants differently, and the Social, Behavioral, and Economic Sciences (SES) directorate is a frequent category of NSF funding in economics. This exclusion tests whether effects are driven by direct financial incentives from existing grants versus broader anticipatory self-censorship. The dependent variable is binary, indicating whether the paper contains any censored word in a Gender, Race, or Environment context. Upper panel shows paper-level estimates with source fixed effects (robust standard errors). Lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Linear probability model estimates. Tabular versions are in Appendix Tables A17 and A18.

significant differences across any dimension: the effect appears broad-based.

By seniority, both senior authors (more than 12 years since PhD, above the author-paper-level median) and junior authors (at or below the median) display negative coefficients of similar magnitude. Two competing mechanisms could produce heterogeneity by career stage. On the one hand, untenured researchers face direct career vulnerability – terminated NSF grants can result in the loss of postdoctoral and research staff positions, and junior faculty depend on grants for tenure cases (e.g., Kozlov et al., 2026). On the other hand, senior faculty carry greater institutional visibility and may internalize university-wide financial pressures: proposed cuts to the NSF budget and attempts to cap indirect cost reimbursements affect not just individual grants but the broader research infrastructure that senior faculty helped build and administer. The absence of a seniority gradient in our data is consistent with both forces operating simultaneously. It is also important to note that our sample consists of NBER and CEPR affiliates, and researchers who co-author with them – an elite demographic in which at least one author on each paper holds a competitive research network affiliation. Junior researchers in this sample are typically not job market candidates but rather early-career faculty who may have substantial outside options, including international mobility, which could attenuate their sensitivity to domestic funding pressure. Moreover, single-authored papers – which are particularly important for junior economists' career advancement – are likely underrepresented in our sample, since arguably few researchers hold an NBER or CEPR affiliation at the time they write their job market paper.

Next we investigate heterogeneous effects by predicted ethnicity – based on US Census 2010 surname data, which assigns each author's last name to its highest-probability racial category. The coefficient of interest is negative and comparable to the main effect among authors with predicted-White surnames and among authors whose surnames are not in the Census file (names that are neither typical English-language nor typical US minority surnames, likely reflecting international backgrounds – e.g., Stantcheva, Acemoglu). Among authors with predicted-Minority surnames (Asian/Pacific Islander, Hispanic, Black), the point estimate is near zero, though the subsample is substantially smaller (N = 3,612 versus 8,946 for predicted-White) and the confidence interval is wide. For authors with non-Census surnames, two competing forces may offset each other: on the one hand, researchers with international backgrounds may face additional vulnerability through visa-related pressures and uncertainty about their immigration status; on the other hand, they are likely competitive on the international academic market and may feature greater geographic mobility, reducing their sensitivity to US-specific funding pressures.

When studying heterogeneous effects by gender, the point estimate for female authors is larger in magnitude than for male authors, but the confidence intervals overlap substantially. In the pre-treatment period, female authors are approximately 8.5 percentage points more likely to use GRE terms; ethnicity minorities are slightly less likely; and there are no differences in the likelihood of using GRE terms across seniority groups. None of the subsample differences are statistically significant, indicating that self-censorship responses are broad-based rather than concentrated in any particular demographic group.[14]

Research output. In what follows we study whether the aforementioned changes in research content may reflect shifts in research productivity. Appendix Figure A13 reports PPML estimates for paper length (number of pages) across the same reference group comparisons: coefficients are close to zero and statistically insignificant at both the paper and author levels. We would expect nonzero estimates if authors systematically shortened their papers – for instance by dropping sections such as demographic heterogeneity analyses.

Appendix Figure A14 estimates the effect on the average number of papers per author per year, comparing each author's pre-period annual average (2020–2024) to their post-period output (2025): point estimates are negative but statistically insignificant and comparable in magnitude to the placebo comparisons, providing no evidence that the content adjustments

[14] The seniority cutoff at 12 years corresponds to the author-paper-level median. Alternative cutoffs (e.g., 6 or 7 years, more aligned with tenure timelines) yield qualitatively similar results. Authors whose PhD graduation year could not be found are assigned zero years since PhD and thus classified as junior. Anecdotal checks suggest these tend to be junior researchers without a personal website yet or industry collaborators; they typically have few papers in the sample.

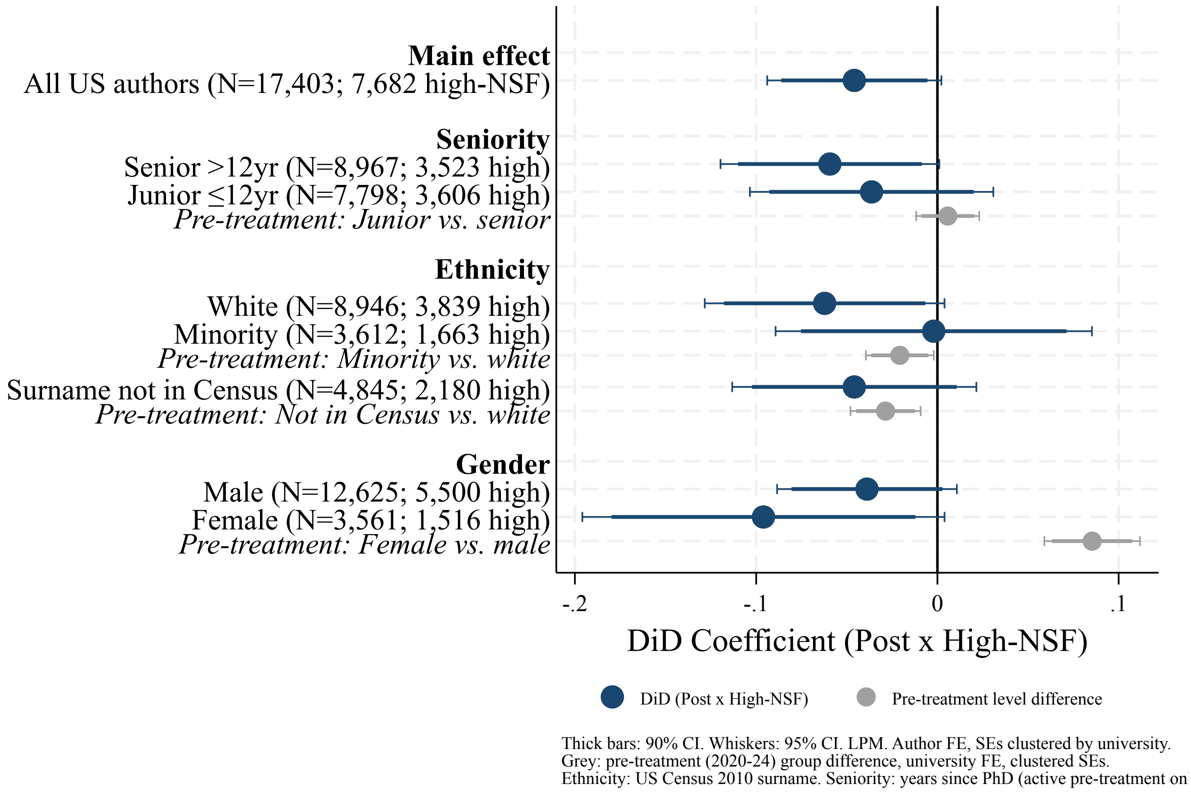


Figure 6: Heterogeneity: Author-Level GRE Content

Notes: This figure reports difference-in-differences coefficients (Post $\times$ High-NSF) estimated separately for different subsamples of US authors. The dependent variable is binary, indicating whether the paper contains any targeted term in a Gender, Race, or Environment context. All specifications include author fixed effects and cluster standard errors by university (LPM). The top row shows the main effect for all US authors; subsequent rows show subsample estimates by seniority (senior >12 years since PhD vs. junior $\leq$12 years), predicted ethnicity (White, Minority, Surname not in Census), and gender (Male, Female). Grey dots show pre-treatment (2020–2024) level differences between groups, estimated with university fixed effects and clustered standard errors: Junior vs. Senior, Minority vs. White, Surname not in Census vs. White, and Female vs. Male. Seniority is computed from PhD graduation years collected via web searches and is time-varying (recomputed for each paper's publication year); the sample is restricted to authors active in the pre-treatment period. Ethnicity is predicted from US Census 2010 surname data ($\sim$162K surnames); Minority groups considered are Asian/Pacific Islander, Hispanic, Black, and American Indian. Gender is classified from first names using Claude Haiku. Thick bars: 90% CI; whiskers: 95% CI. Tabular version is in Appendix Table A19.

documented above led to significant delays or reductions in research output in 2025.

5 Conclusion

Self-censorship in science is particularly hard to measure. In February 2025, the US administration began scrutinizing federally funded research based on a list of words to be avoided. One might have expected researchers to ignore this change and working solely according to the principles of academic freedom. Even if researchers reacted, one might have expected the impact to be limited to the content of new NSF grant applications. Instead, we find that economists at universities heavily dependent on NSF funding changed the content of their ongoing research almost immediately by reducing the use of newly banned words in sensitive contexts. This impact is not narrowly driven by research that is directly funded by the NSF. Instead, the changing funding environment appears to have triggered a broader effect on the choice of research topics. As academic freedom has declined around the world in recent years, strong self-censorship responses have wide-ranging implications for the resilience of science.

Understanding better how scholars react to political pressure in various contexts is important, and future work on this is encouraged. Recent research has shown that, on average, populist governments are followed after 15 years by a GDP per capita that is 10 percent lower than a plausible non-populist counterfactual (Funke et al., 2023). As such an economic record could backfire at the polls, populist leaders have powerful incentives to not only weaken democratic institutions, but also silence independent scrutiny by the media and science. To defend civil and political rights and to support evidence-based policymaking, strengthening academic freedom is key.

References

Alabrese, E. (2022). Bad Science: Retractions and Media Coverage. Available at SSRN: `https://ssrn.com/abstract=4324218`.

Alabrese, E., F. Capozza, and P. Garg (2024). Politicized Scientists: Credibility Cost of Political Expression on Twitter. Available at SSRN: `https://ssrn.com/abstract=4922427`.

Ash, E., D. L. Chen, and S. Naidu (2025). Ideas Have Consequences: The Impact of Law and Economics on American Justice. *The Quarterly Journal of Economics*. Advance access.

Banerjee, A. V., S. Chassang, and E. Snowberg (2017). Decision Theoretic Approaches to Experiment Design and External Validity. In *Handbook of economic field experiments*, Volume 1, pp. 141–174. Elsevier.

Borjas, G. J. and N. Breznau (2026). Ideological Bias in the Production of Research Findings. *Science Advances 12*(1), eadz7173.

Brodeur, A., N. Cook, and A. Heyes (2020). Methods Matter: P-hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review 110*(11), 3634–3660.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics 8*(1), 1–32.

Chassang, S., G. Padró i Miquel, and E. Snowberg (2012). Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *American Economic Review 102*(4), 1279–1309.

Clark, C. J., L. Jussim, K. Frey, S. T. Stevens, M. al Gharbi, K. Aquino, J. M. Bailey, N. Barbaro, R. F. Baumeister, A. Bleske-Rechek, D. Buss, S. Ceci, M. Del Giudice, P. H. Ditto, J. P. Forgas, D. C. Geary, G. Geher, S. Haider, N. Honeycutt, H. Joshi, A. I. Krylov, E. Loftus, G. Loury, L. Lu, M. Macy, C. C. Martin, J. McWhorter, G. Miller, P. Paresky, S. Pinker, W. Reilly, C. Salmon, S. Stewart-Williams, P. E. Tetlock, W. M.

Williams, A. E. Wilson, B. M. Winegard, G. Yancey, and W. von Hippel (2023). Prosocial Motives Underlie Scientific Censorship by Scientists: A Perspective and Research Agenda. *Proceedings of the National Academy of Sciences 120*(48), e2301642120.

Fourcade, M., E. Ollion, and Y. Algan (2015). The Superiority of Economists. *Journal of economic perspectives 29*(1), 89–114.

Funke, M., M. Schularick, and C. Trebesch (2023). Populist Leaders and the Economy. *American Economic Review 113*(12), 3249–3288.

Furnas, A. C., R. Jia, M. E. Roberts, and D. Wang (2026). Geopolitics in the Evaluation of International Scientific Collaboration. Working Paper 34789, National Bureau of Economic Research.

Garg, P. and T. Fetzer (2025a). Causal Claims in Economics. *Working Paper*.

Garg, P. and T. Fetzer (2025b). Political Expression of Academics on Twitter. *Nature Human Behaviour*, 1815.

Grosjean, P., F. Masera, and H. Yousaf (2023). Inflammatory Political Campaigns and Racial Bias in Policing. *The Quarterly Journal of Economics 138*(1), 413–463.

Guriev, S. and D. Treisman (2019). Informational Autocrats. *Journal of Economic Perspectives 33*(4), 100–127.

Hamermesh, D. S. (2013). Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature 51*(1), 162–72.

Havel, V. (2009). *The Power of the Powerless (Routledge Revivals): Citizens against the State in Central-Eastern Europe*. Routledge.

Hjort, J., D. Moreira, G. Rao, and J. F. Santini (2021). How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities. *American Economic Review 111*(5), 1442–1480.

Honeycutt, N., S. T. Stevens, and E. Kaufmann (2023). The Academic Mind in 2022: What Faculty Think About Free Expression and Academic Freedom on Campus. Report, The Foundation for Individual Rights and Expression.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics 133*(2), 927–991.

Ioannidis, J. P., T. D. Stanley, and H. Doucouliagos (2017). The Power of Bias in Economics Research.

Jelveh, Z., B. Kogut, and S. Naidu (2024). Political Language in Economics. *The Economic Journal 134*(662), 2439–2469.

Kasy, M. (2021). Of Forking Paths and Tied Hands: Selective Publication of Findings, and what Economists should do about it. *Journal of Economic Perspectives 35*(3), 175–192.

Kinzelbach, K., S. I. Lindberg, L. Lott, and A. V. Panaro (2025). Academic Freedom Index–Update 2025. *Available at SSRN 5651510*.

Kozlov, M., J. Tollefson, and D. Garisto (2026). US Science after a Year of Trump: What Has Been Lost and What Remains. Nature.com, accessed: 2026-01-30.

Ratcliff, C. L., B. Harvill, and R. Wicke (2023). Understanding Public Preferences for Learning about Uncertain Science: Measurement and Individual Difference Correlates. *Frontiers in Communication Volume 8 - 2023*.

Rubin, A. and E. Rubin (2021). Systematic Bias in the Progress of Research. *Journal of Political Economy 129*(9), 2666–2719.

Vanden Eynde, O., P. M. Kuhn, and A. Moradi (2018). Trickle-Down Ethnic Politics: Drunk and Absent in the Kenya Police Force (1957-1970). *American Economic Journal: Economic Policy 10*(3), 388–417.

Waldinger, F. (2012). Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *The Review of Economic Studies 79*(2), 838–861.

Widmer, P. (2024). Is Propaganda Front-Page News? *Working Paper*.

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Evidence from the Rwandan Genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

Yourish, K., A. Daniel, S. Datar, I. White, and L. Gamio (2025). These Words Are Disappearing in the New Trump Administration. The New York Times. 7 March 2025.

A Data Appendix

A.1 Context Classification and Full List of Banned Words

In this Appendix subsection we will start by discussing the prompt used for the classification of contexts and then include below a list of all banned terms (Table A1).

For each word-paper pair (i.e., each occurrence of a targeted term in an abstract), we classify the context in which the word is used. The following prompt is sent to Claude Haiku (model `claude-haiku-4-5-20251001`), with {`word`} and {`abstract`} filled for each observation:

> *The following word has been censored by the Trump administration: {word}*
>
> *This word appears in the economics research abstract below. In which context is this word used?*
>
> *Contexts:*
>
> \- *Economic inequality*
>
> \- *Race*
>
> \- *Gender*
>
> \- *Environment*
>
> \- *Cannot say or other*
>
> *If multiple contexts could apply, choose the dominant one.*
>
> *Abstract: "{abstract}"*
>
> *Return ONLY the context name.*

A.2 Gender Classification

Using Claude Haiku (model `claude-haiku-4-5-20251001`), we classify author gender from first names. The following prompt is used:

Classify the likely gender of this person based on their name.

Respond with exactly one of: male, mostly male, mostly female, female, impossible to know.

Name: {name}

Responses are collapsed into a binary variable: "male" and "mostly male" map to male; "female" and "mostly female" map to female; "impossible to know" is treated as missing.

A.3 Ethnicity Prediction

Predicted ethnicity is based on the US Census Bureau's 2010 Frequently Occurring Surnames file, which lists approximately 162,000 surnames that together cover the vast majority of the US population. For each surname, the file provides the percentage distribution across six racial/ethnic categories: White, Black, Asian/Pacific Islander, Hispanic, American Indian/Alaska Native, and Two or More Races. We extract each author's last name, normalize it to uppercase ASCII, and look it up in the Census file. The predicted ethnicity is the category with the highest probability for that surname. Authors whose surnames do not appear in the file are classified as "Unknown."

A.4 PhD Graduation Years

PhD graduation years are collected via automated web searches using Claude Haiku agents, which query university faculty pages, CVs, Wikipedia, RePEc profiles, and Google Scholar for each author. Results are manually verified. Seniority (years since PhD) is computed as the difference between the paper's publication year and the PhD graduation year and is therefore time-varying.

Table A1: List of Targeted Terms

accessible, activism, activists, advocacy, advocate, advocates, affirming care, all-inclusive, allyship, anti-racism, antiracist, assigned at birth, assigned female at birth, assigned male at birth, at risk, barrier, barriers, belong, bias, biased, biased toward, biases, biases towards, biologically female, biologically male, BIPOC, Black, breastfeed + people, breastfeed + person, chestfeed + people, chestfeed + person, clean energy, climate crisis, climate science, commercial sex worker, community diversity, community equity, confirmation bias, cultural competence, cultural differences, cultural heritage, cultural sensitivity, culturally appropriate, culturally responsive, DEI, DEIA, DEIAB, DEIJ, disabilities, disability, discriminated, discrimination, discriminatory, disparity, diverse, diverse backgrounds, diverse communities, diverse community, diverse group, diverse groups, diversified, diversify, diversifying, diversity, enhance the diversity, enhancing diversity, environmental quality, equal opportunity, equality, equitable, equitableness, equity, ethnicity, excluded, exclusion, expression, female, females, feminism, fostering inclusivity, GBV, gender, gender based, gender based violence, gender diversity, gender identity, gender ideology, gender-affirming care, genders, Gulf of Mexico, hate speech, health disparity, health equity, hispanic minority, historically, identity, immigrants, implicit bias, implicit biases, inclusion, inclusive, inclusive leadership, inclusiveness, inclusivity, increase diversity, increase the diversity, indigenous community, inequalities, inequality, inequitable, inequities, inequity, injustice, institutional, intersectional, intersectionality, key groups, key people, key populations, Latinx, LGBT, LGBTQ, marginalize, marginalized, men who have sex with men, mental health, minorities, minority, most risk, MSM, multicultural, Mx, Native American, non-binary, nonbinary, oppression, oppressive, orientation, people + uterus, people-centered care, person-centered, person-centered care, polarization, political, pollution, pregnant people, pregnant person, pregnant persons, prejudice, privilege, privileges, promote diversity, promoting diversity, pronoun, pronouns, prostitute, race, race and ethnicity, racial, racial diversity, racial identity, racial inequality, racial justice, racially, racism, segregation, sense of belonging, sex, sexual preferences, sexuality, social justice, sociocultural, socioeconomic, status, stereotype, stereotypes, systemic, systemically, they/them, trans, transgender, transsexual, trauma, traumatic, tribal, unconscious bias, underappreciated, underprivileged, underrepresentation, underrepresented, underserved, undervalued, victim, victims, vulnerable populations, women, women and underrepresented

Notes: 197 terms, based on the 174 terms reported in Yourish et al. (2025) plus morphological variants. The "+" notation denotes word combinations (e.g., "breastfeed + people" matches "breastfeeding people"). For text matching, we expand the list to include UK spellings (e.g., marginalise, polarisation) and hyphenation variants (e.g., anti-racism, antiracism), yielding 224 search patterns in total.

Table A2: Summary Statistics: Paper Level

	All	US High-NSF	US Low-NSF
Panel A: Paper characteristics			
Number of papers	14,412	1,623	8,407
Targeted word count (abstract)	1.111	1.172	1.139
	(2.247)	(2.256)	(2.325)
Targeted word count excl. political (abstract)	0.907	0.982	0.936
	(2.084)	(2.105)	(2.166)
Targeted word count (full text)	82.53	83.13	87.61
	(124.77)	(125.00)	(133.82)
Social-sense word count (abstract)	0.627	0.688	0.654
	(1.851)	(1.927)	(1.919)
Number of pages	56.7	54.8	58.3
	(23.5)	(23.3)	(24.2)
Number of authors	3.06	2.83	3.23
	(1.51)	(1.26)	(1.65)
Panel B: Topic classification			
Gender, race, or environment	0.138	0.153	0.144
Gender	0.075	0.063	0.077
Race	0.036	0.055	0.044
Environment	0.036	0.053	0.032
Panel C: Funding and affiliation			
Number of US-affiliated authors	1.45	2.55	2.00
Number of UK-affiliated authors	0.20	0.02	0.14
Share NBER	0.556	0.839	0.731
Share CEPR	0.444	0.161	0.269
Any author at red-state university	0.181	0.121	0.192

Notes: Standard deviations in parentheses. US High-NSF = papers where the majority of US-affiliated authors are from universities with above-median NSF funding share. US Low-NSF = US-affiliated papers not classified as High-NSF. Topic shares are computed over all papers (including those with zero targeted words). Red-state = university located in a state that voted Republican in the 2024 presidential election.

Table A3: Summary Statistics: Author Level

	All US	US High-NSF	US Low-NSF
Panel A: Paper characteristics			
Number of author-paper pairs	20,893	7,682	13,211
Unique authors	6,827	2,576	4,463
GRE content	0.154	0.156	0.153
Targeted word count (abstract)	1.178	1.193	1.170
	(2.347)	(2.352)	(2.344)
Panel B: Author characteristics			
Female	0.232	0.216	0.241
White	0.511	0.500	0.518
Minority	0.222	0.216	0.226
Surname not in Census	0.266	0.284	0.256
Years since PhD	14.7	14.3	14.9
	(13.0)	(12.4)	(13.3)
Senior (>12 years since PhD)	0.475	0.472	0.477
Red state	0.239	0.168	0.281

Notes: Unit of observation is the author-paper pair, restricted to US-affiliated authors matched to HERD funding data. Standard deviations in parentheses. US High-NSF = author at a US university with above-median NSF funding share. US Low-NSF = US-affiliated author not classified as High-NSF. Ethnicity predicted from US Census 2010 surname data. Senior = more than 12 years since PhD (author-paper-level median). Red state = university in a state that voted Republican in the 2024 presidential election.

A.5 Descriptive Summary Statistics

In this subsection, we will present tables of descriptive summary statistics (Tables A2 and A3), followed by topic-specific word clouds (Figure A1).

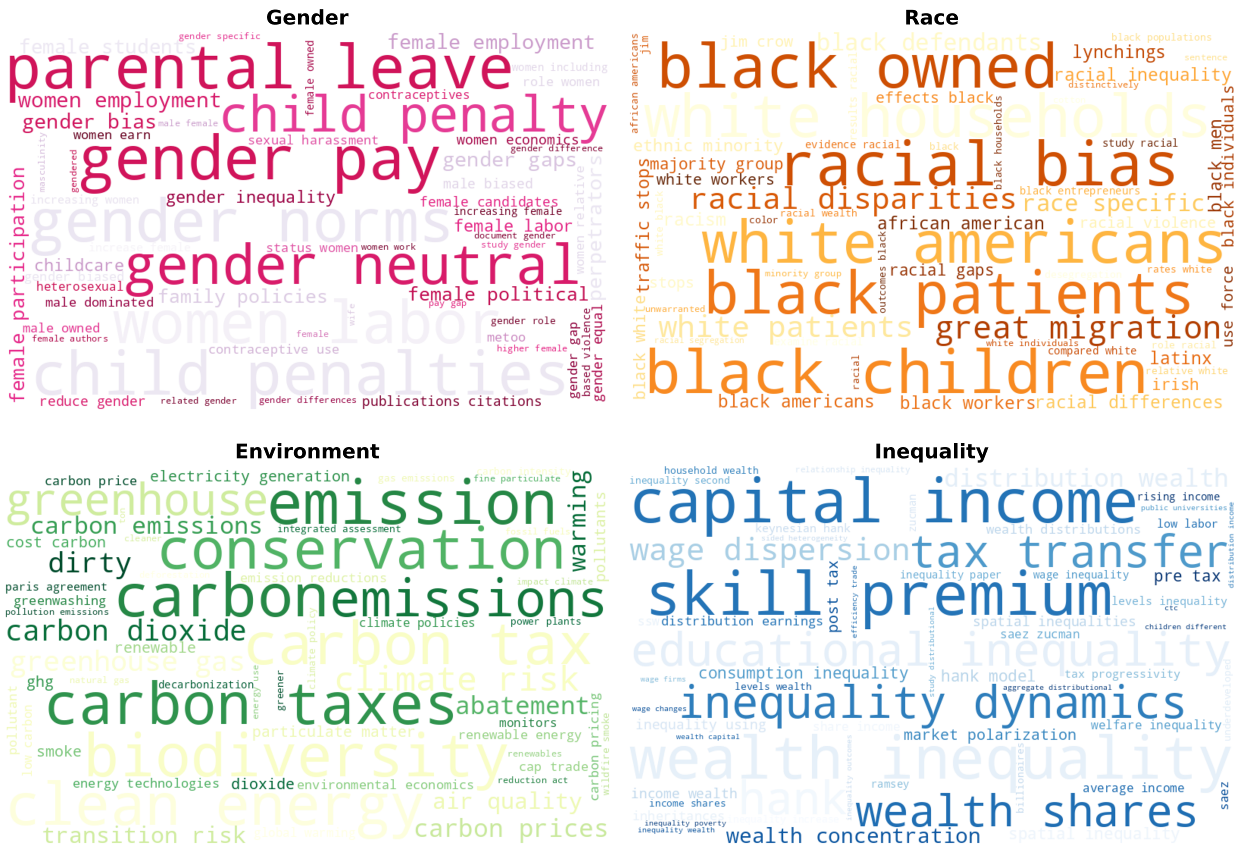


Figure A1: Odds-Ratio Wordclouds by Topic Category

Notes: Each panel displays the most distinctive unigrams and bigrams for abstracts classified in the given category, with word size proportional to the log odds ratio of appearing in abstracts of that category relative to abstracts not in that category. Abstracts are tokenized into unigrams and bigrams (minimum document frequency of 5, English stop words excluded). Based on 5,477 classified abstracts from NBER and CEPR working papers (2020–2025). Classification by Claude Haiku based on surrounding text in the abstract.

B Results Appendix

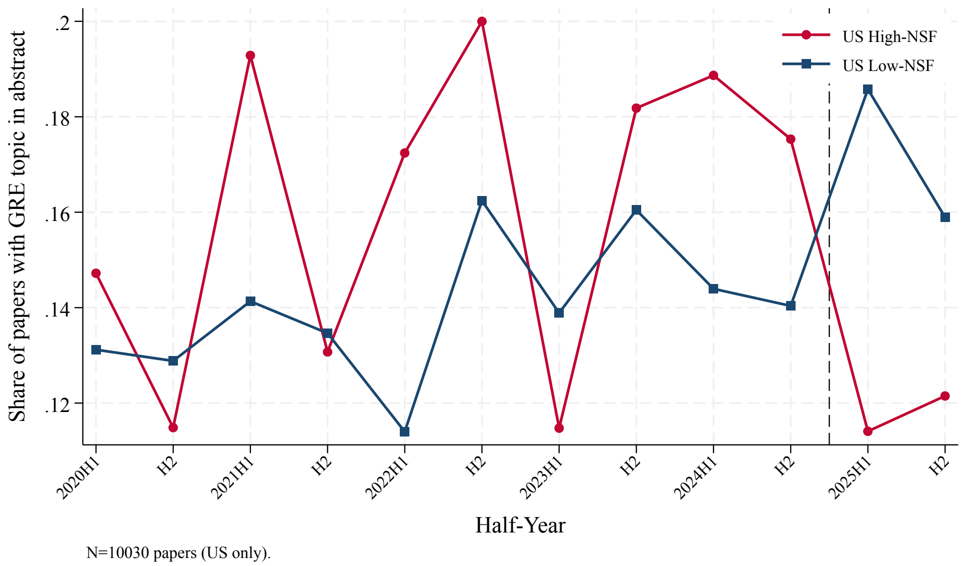
We begin with a descriptive look at the raw data. Figure A2 plots the share of papers containing GRE-related content by half-year, separately for US High-NSF and Low-NSF institutions, alongside the gap between the two groups. The two groups follow similar trends throughout the pre-treatment period, before diverging sharply after the policy change.

We then present robustness checks for both the event study (Figure 2) and reference group comparisons (Figure 3). While the first set of robustness figures will focus on event studies (Appendix Figures A3–A6), the follow series of robustness figures will focus on coefficient plots for reference group comparisons and A7–A12).

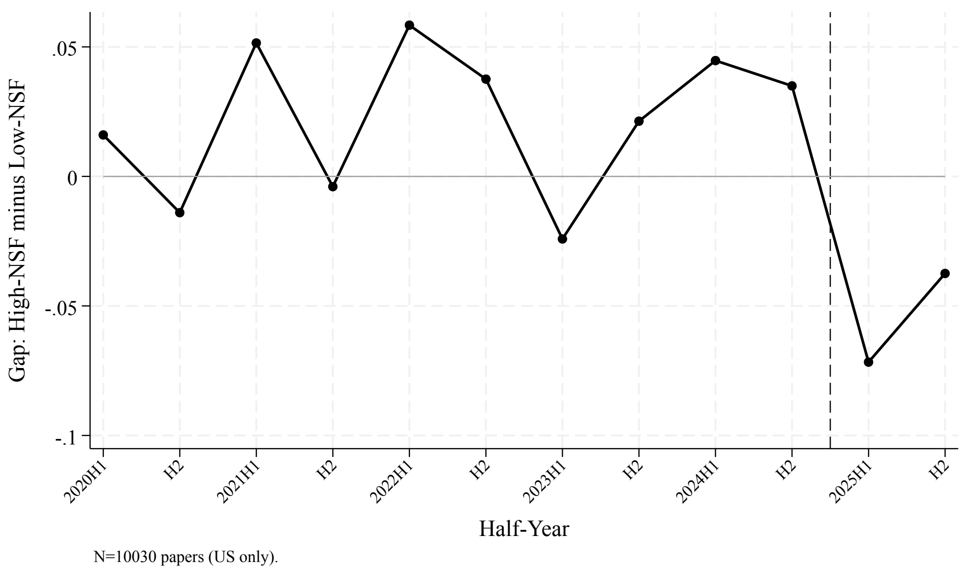
We vary the outcome measure: GRE word counts in the abstract (Figures A3 and A7), any censored word in the abstract regardless of context (Figures A4 and A8), Qwen3-32B full-text GRE word counts (Figures A5 and A9), and raw censored word counts in the full text (Figures A6 and A10). We also report reference group comparisons using only 2024 as the pre-treatment period (Figure A11) and controlling for Red/Blue state $\times$ Post (Figure A12).

Concerning mechanisms, we present estimates for paper length (Figure A13) and paper counts (Figure A14).

Finally, we include a table for each coefficient plot in the main text or Appendix (see Appendix Tables A4-A24).



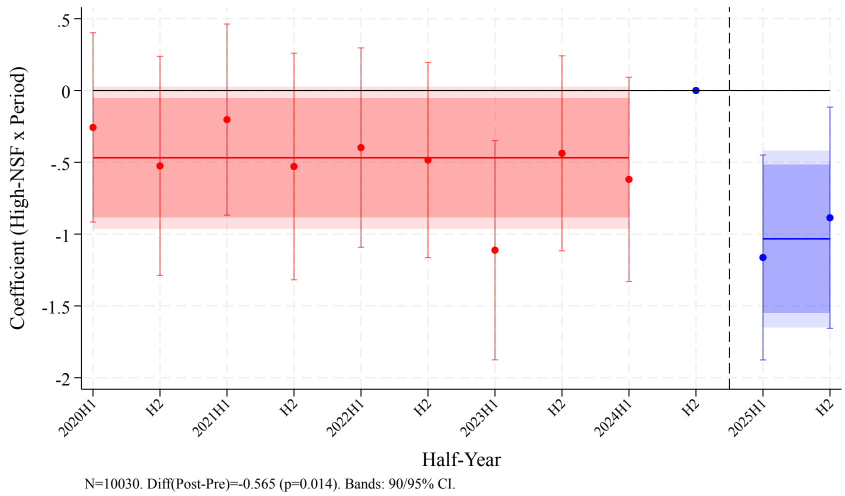
(a) Raw Shares by Group



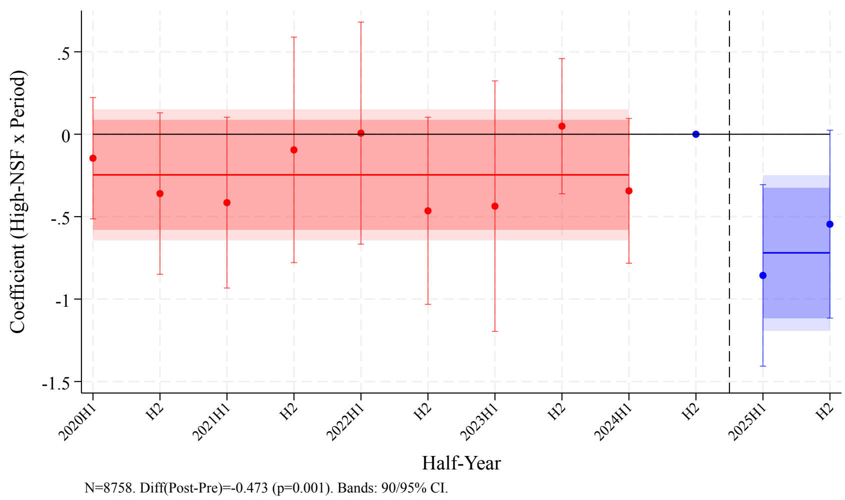
(b) Gap (High-NSF minus Low-NSF)

Figure A2: Raw GRE Content Shares: High-NSF vs Low-NSF Universities

Notes: Upper panel plots the share of papers containing at least one targeted term in a Gender, Race, or Environment context, separately for US High-NSF and Low-NSF institutions, by half-year. Lower panel plots the difference (High-NSF minus Low-NSF). The dashed vertical line indicates the Trump inauguration (January 2025). US papers only.



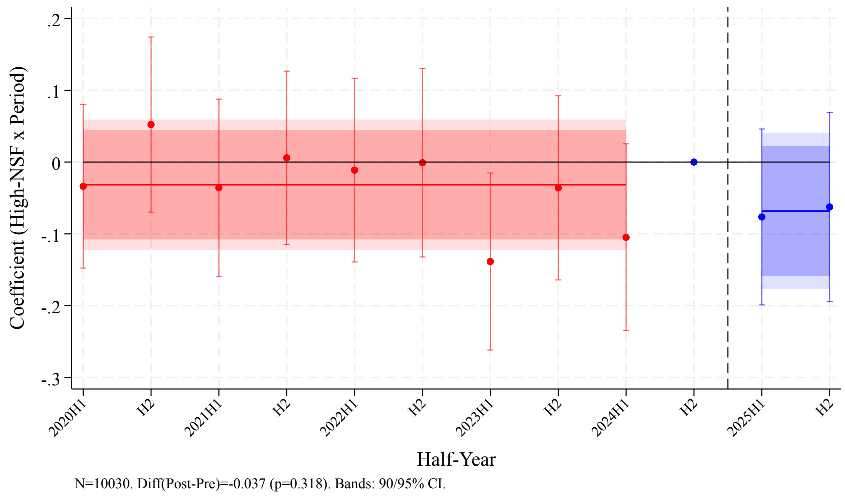
(a) Paper-Level: GRE Word Count in Abstract (Reference: 2024H2)



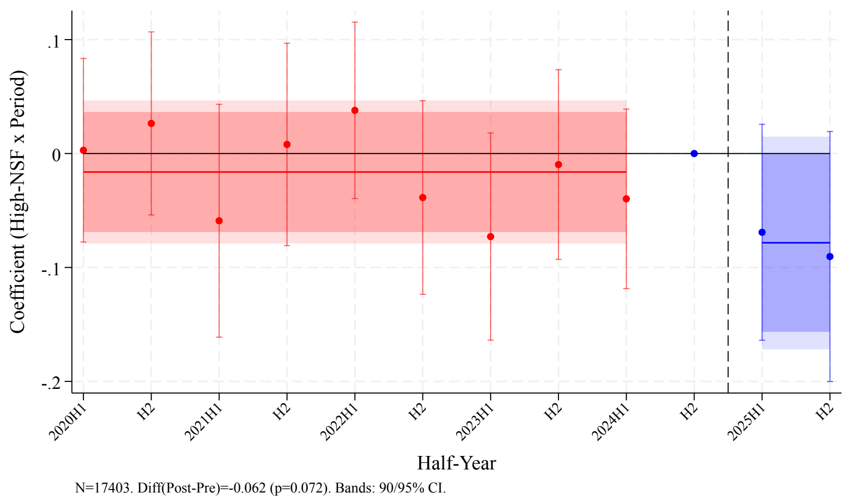
(b) Author-Level: GRE Word Count in Abstract (Reference: 2024H2)

Figure A3: Event Study Robustness: GRE Word Count in Abstract

Notes: This figure replicates Figure 2 using the count of words classified into Gender, Race, or Environment categories in the abstract (rather than a binary indicator). Half-yearly difference-in-differences coefficients comparing US High-NSF institutions to US Low-NSF institutions, with 2024H2 as the reference period. Red shading indicates pre-treatment periods; blue shading indicates post-treatment periods. Darker bands represent 90% confidence intervals; lighter bands represent 95% confidence intervals. Horizontal lines show pooled pre- and post-treatment averages. Upper panel shows paper-level estimates with source fixed effects (robust standard errors). Lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). PPML estimates.



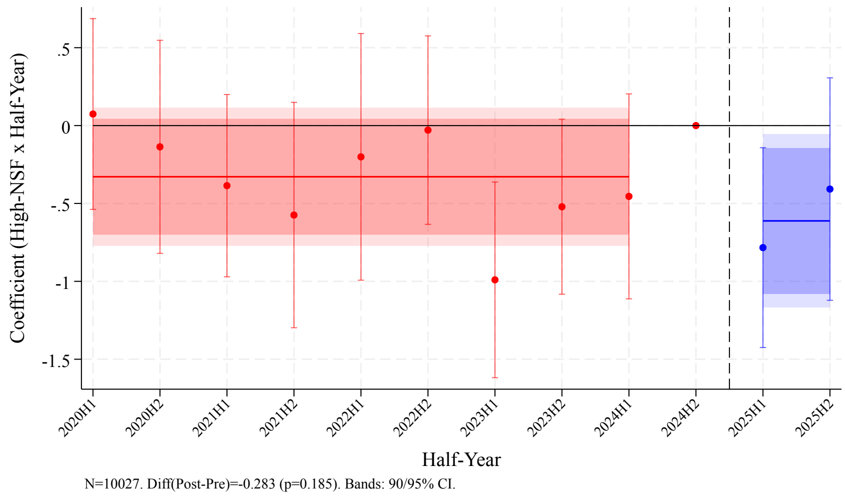
(a) Paper-Level: Any Targeted Word in Abstract (Reference: 2024H2)



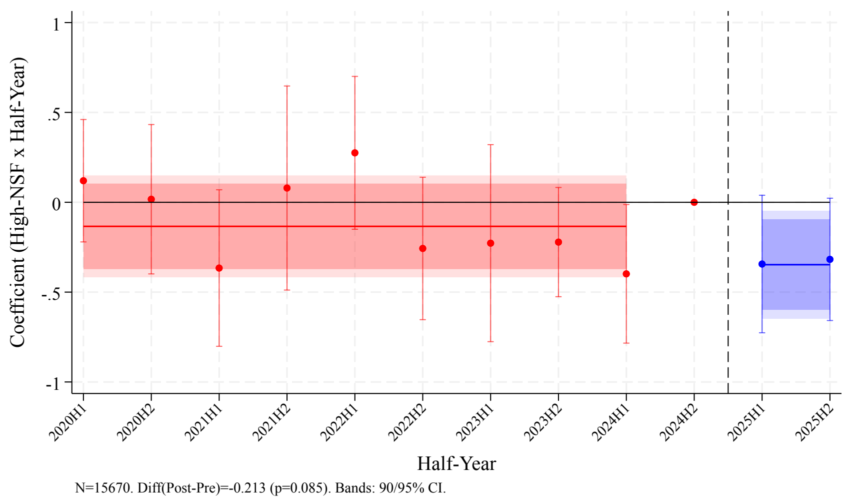
(b) Author-Level: Any Targeted Word in Abstract (Reference: 2024H2)

Figure A4: Event Study Robustness: Any Censored Word in Abstract

Notes: This figure shows half-yearly difference-in-differences coefficients comparing US High-NSF institutions to US Low-NSF institutions, with 2024H2 as the reference period. The dependent variable is binary, indicating whether any censored word appears in the abstract (regardless of context). Red shading indicates pre-treatment periods; blue shading indicates post-treatment periods. Darker bands represent 90% confidence intervals; lighter bands represent 95% confidence intervals. Horizontal lines show pooled pre- and post-treatment averages. Upper panel shows paper-level estimates with source fixed effects (robust standard errors). Lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). Linear probability model estimates.



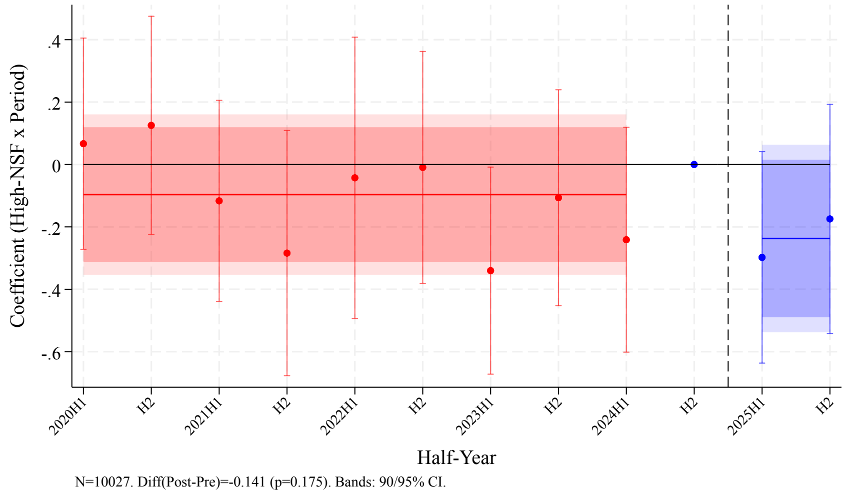
(a) Paper-Level: GRE Word Counts, Qwen3 Full-Text (Reference: 2024H2)



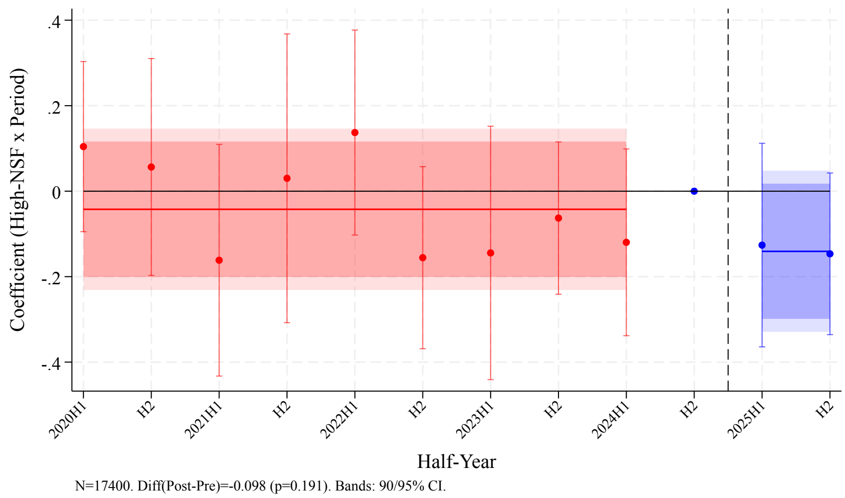
(b) Author-Level: GRE Word Counts, Qwen3 Full-Text (Reference: 2024H2)

Figure A5: Event Study Robustness: Qwen3 Full-Text GRE Word Counts

Notes: This figure replicates Figure 2 using an alternative topic measure. Instead of classifying abstracts with Claude Haiku, the dependent variable is the count of words classified into Gender, Race, or Environment categories in the full paper text using Qwen3-32B (via Groq API). Half-yearly difference-in-differences coefficients comparing US High-NSF institutions to US Low-NSF institutions, with 2024H2 as the reference period. Red shading indicates pre-treatment periods; blue shading indicates post-treatment periods. Darker bands represent 90% confidence intervals; lighter bands represent 95% confidence intervals. Horizontal lines show pooled pre- and post-treatment averages. Upper panel shows paper-level estimates with source fixed effects (robust standard errors). Lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). Page count is included as a control variable. PPML estimates.

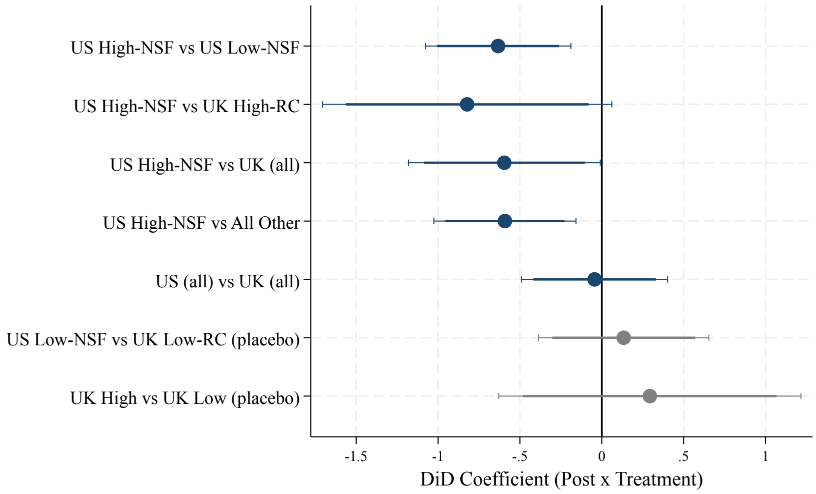


(a) Paper-Level: Full-Text Wordcount (Reference: 2024H2)

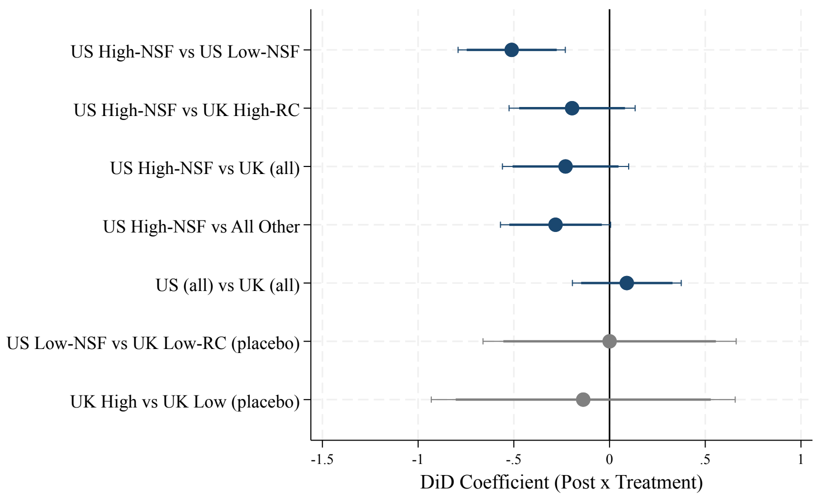


(b) Author-Level: Full-Text Wordcount (Reference: 2024H2)

Figure A6: Event Study Robustness: Half-Yearly Censored Word Counts in Full Text
Notes: This figure replicates Figure 2 using counts of censored words in the full paper text (rather than abstracts only). Half-yearly difference-in-differences coefficients comparing US High-NSF institutions to US Low-NSF institutions, with 2024H2 as the reference period. Red shading indicates pre-treatment periods; blue shading indicates post-treatment periods. Darker bands represent 90% confidence intervals; lighter bands represent 95% confidence intervals. Upper panel shows paper-level estimates with source fixed effects (robust standard errors). Lower panel shows author-level estimates with author fixed effects (standard errors clustered by university). Page count is included as a control variable. PPML estimates.

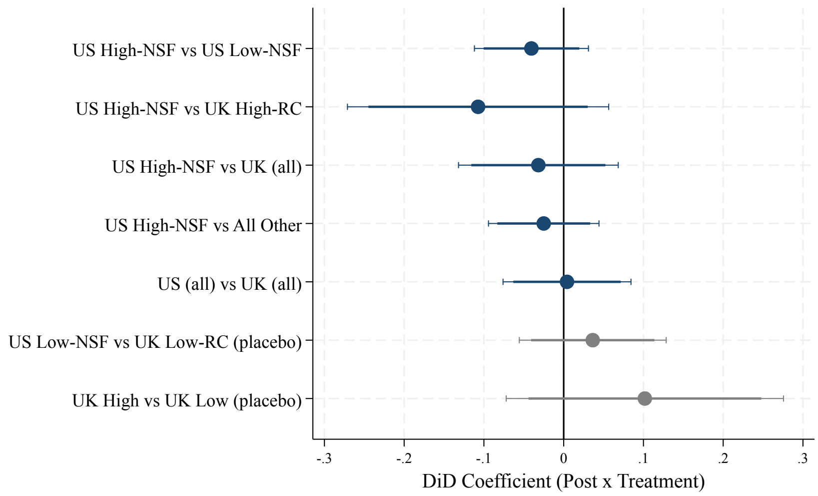


(a) Paper-Level: GRE Word Count in Abstract

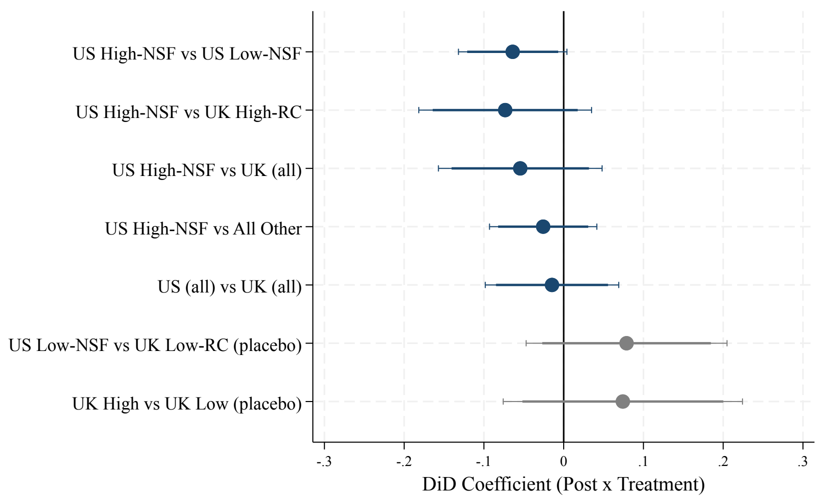


(b) Author-Level: GRE Word Count in Abstract

Figure A7: Reference Group Comparisons Robustness: GRE Word Count in Abstract
Notes: This figure replicates Figure 3 using the count of words classified into Gender, Race, or Environment categories in the abstract (rather than a binary indicator). See Figure 3 notes for description of comparisons. Upper panel shows paper-level estimates; lower panel shows author-level estimates. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. PPML estimates.

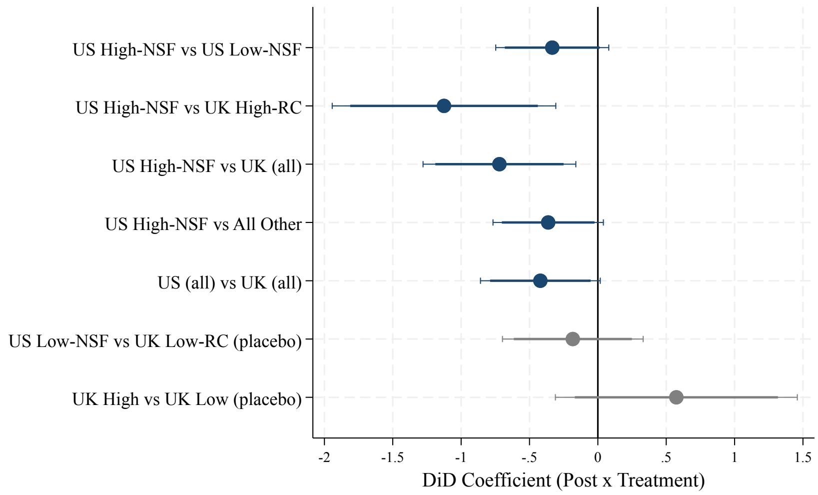


(a) Paper-Level: Any Targeted Word in Abstract

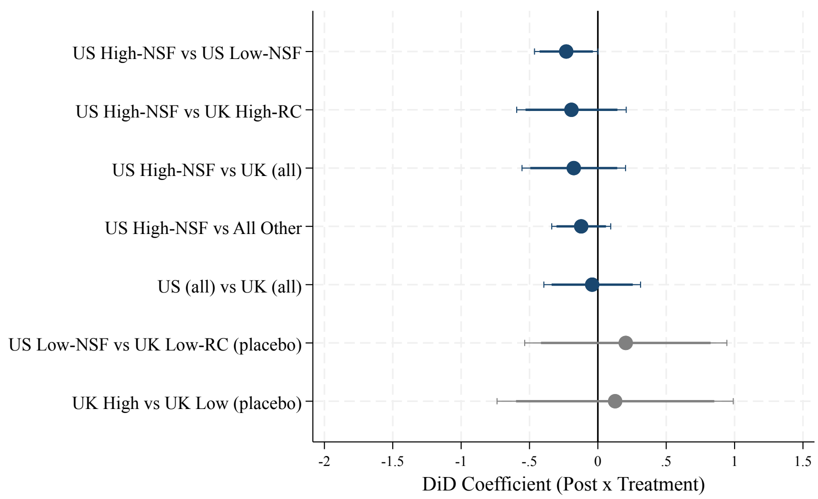


(b) Author-Level: Any Targeted Word in Abstract

Figure A8: Reference Group Comparisons Robustness: Any Targeted Word in Abstract
Notes: This figure replicates Figure 3 using a binary indicator for whether any censored word appears in the abstract (regardless of context). See Figure 3 notes for description of comparisons. Upper panel shows paper-level estimates; lower panel shows author-level estimates. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Linear probability model estimates. Tabular versions are in Appendix Tables A8 and A9.



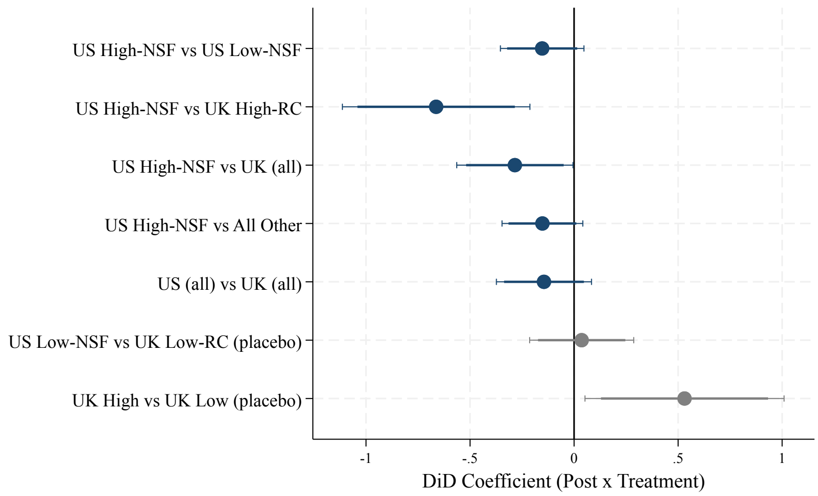
(a) Paper-Level: GRE Word Counts, Qwen3 Full-Text



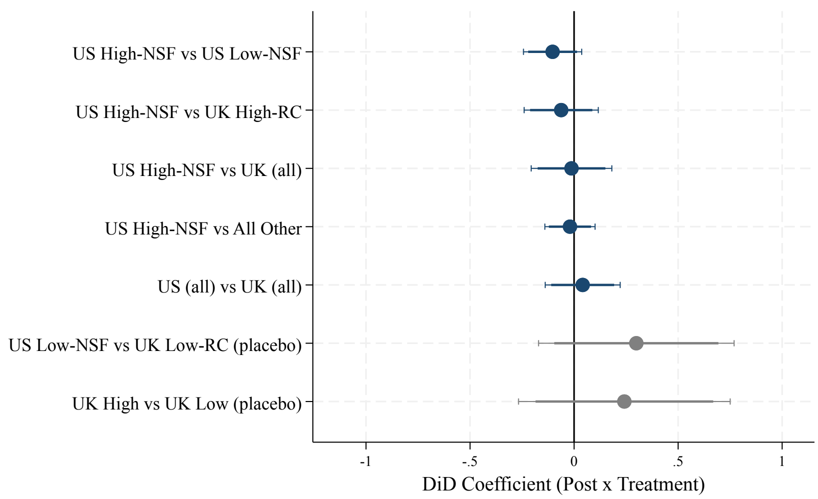
(b) Author-Level: GRE Word Counts, Qwen3 Full-Text

Figure A9: Reference Group Comparisons Robustness: Qwen3 Full-Text GRE Word Counts

Notes: This figure replicates Figure 3 using an alternative topic measure based on full paper text rather than abstracts. The dependent variable is the count of words classified into Gender, Race, or Environment categories in the full paper text using Qwen3-32B (via Groq API). See Figure 3 notes for description of comparisons. Page count is included as a control variable. Upper panel shows paper-level estimates; lower panel shows author-level estimates. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. PPML estimates. Tabular versions are in Appendix Tables A10 and A11.



(a) Paper-Level: Full-Text Wordcount



(b) Author-Level: Full-Text Wordcount

Figure A10: Reference Group Comparisons Robustness: Counts in Full Text

Notes: This figure replicates Figure 3 using counts of censored words in the full paper text. See Figure 3 notes for description of comparisons. Page count is included as a control variable. Upper panel shows paper-level estimates; lower panel shows author-level estimates. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. PPML estimates. Tabular versions are in Appendix Tables A12 and A13.

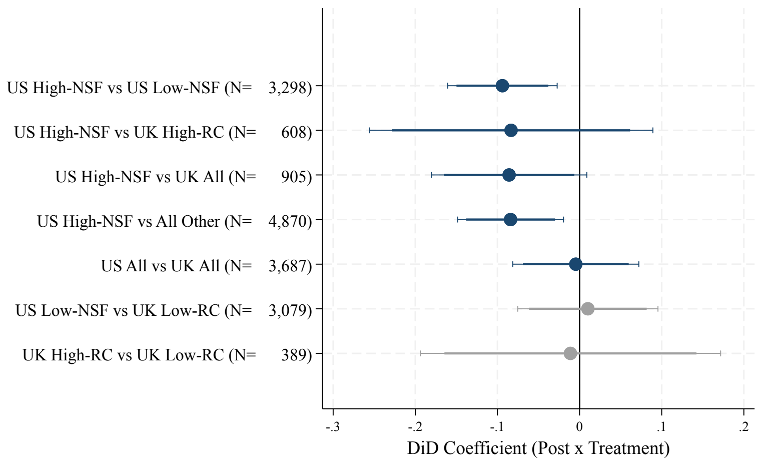
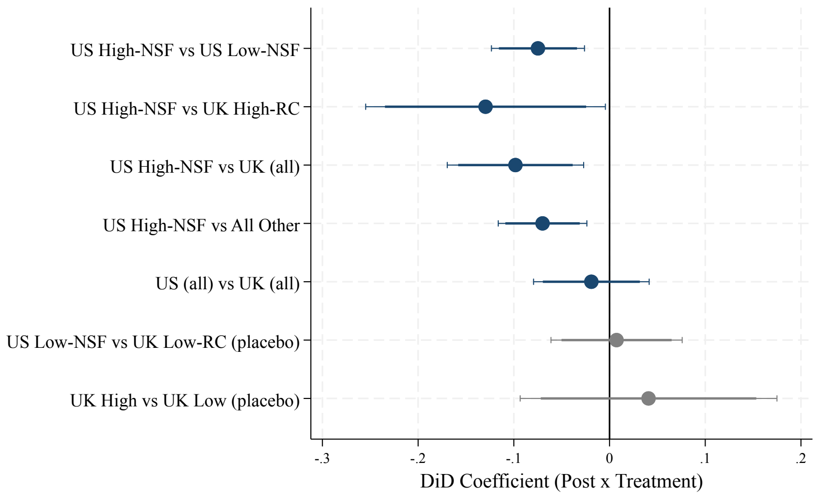
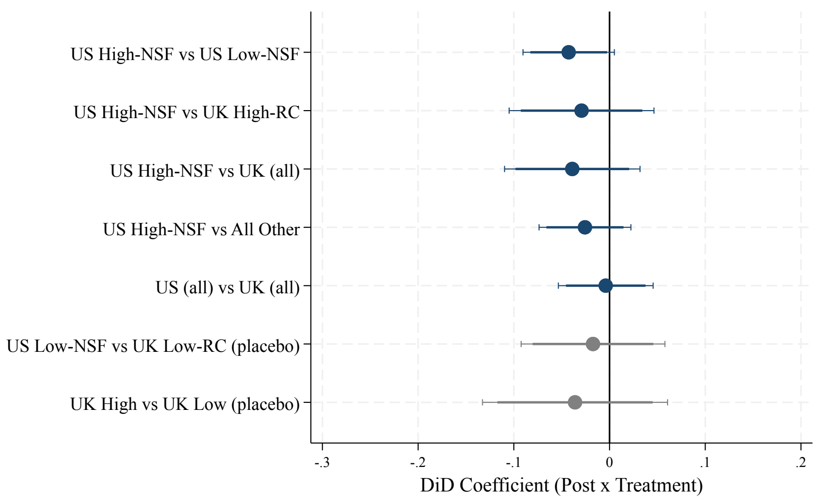


Figure A11: Reference Group Comparisons Robustness: 2024 Only as Pre-Treatment Period (Gender+Race+Environment Content)

Notes: This figure replicates Figure 3 using only 2024 data as the pre-treatment period (rather than 2020–2024). The dependent variable is binary, indicating whether the paper contains any censored word in a Gender, Race, or Environment context. This specification tests robustness to potential pre-trends by using only the most recent pre-treatment year. Only paper-level estimates are shown, as author fixed effects require within-author variation across both periods, leaving too few observations for reliable inference with only one pre-treatment year. See Figure 3 notes for description of comparisons. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Linear probability model estimates. Robust standard errors.



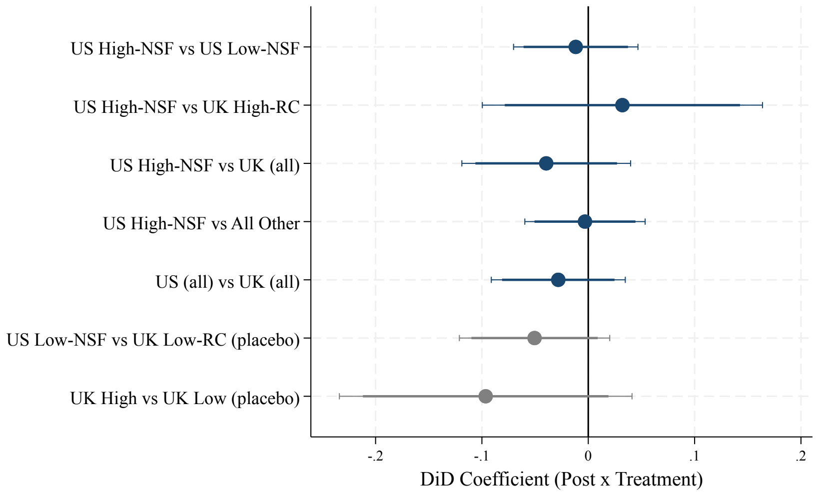
(a) Paper-Level: GRE Content (controlling for Red/Blue x Post)



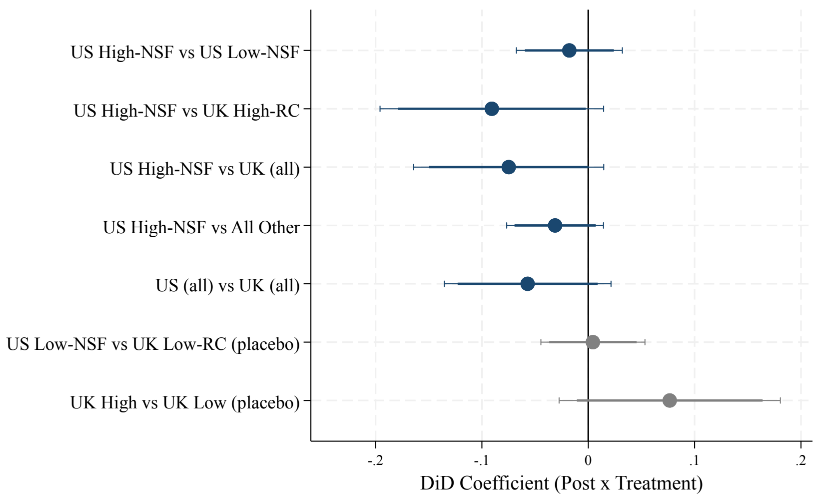
(b) Author-Level: GRE Content (controlling for Red/Blue x Post)

Figure A12: Reference Group Comparisons Robustness: Controlling for Red/Blue State × Post

Notes: This figure replicates Figure 3 while adding Red State × Post as an additional control variable, where Red State indicates universities located in states that voted Republican in the 2024 presidential election. This specification absorbs any differential trends in research content between universities in Republican- and Democratic-leaning states, ensuring that the NSF exposure effect is not confounded by state-level political climate. Upper panel shows paper-level estimates with robust standard errors; lower panel shows author-level estimates with author fixed effects and standard errors clustered by university. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. Linear probability model estimates. Tabular versions are in Appendix Tables A20 and A21.



(a) Paper-Level: Paper Length (Pages)



(b) Author-Level: Paper Length

Figure A13: Reference Group Comparisons: Paper Length (Number of Pages)

Notes: This figure reports pooled difference-in-differences estimates with the number of pages as the dependent variable. See Figure 3 notes for description of comparisons. Upper panel shows paper-level estimates; lower panel shows author-level estimates. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. PPML estimates. Paper level: robust standard errors. Author level: author fixed effects, standard errors clustered by university. Grey = placebo comparisons. Tabular versions are in Appendix Tables A22 and A23.

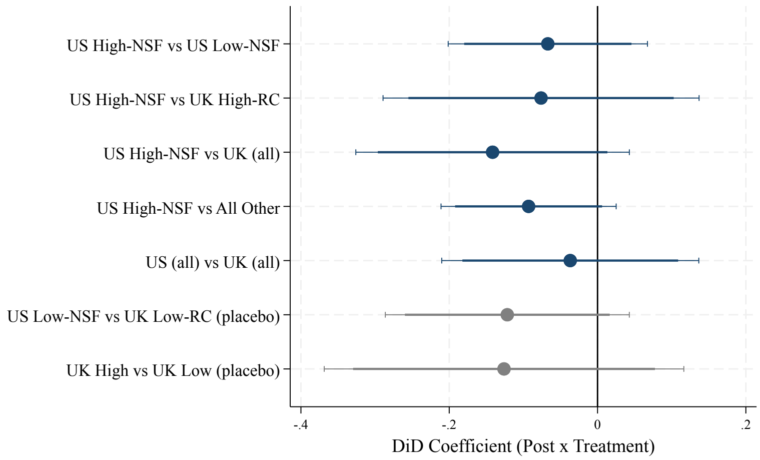


Figure A14: Reference Group Comparisons: Paper Counts (Number of Papers per Author per Year)

Notes: This figure reports pooled difference-in-differences estimates with the average number of papers per author per year as the dependent variable. For each author, we compute the average annual paper count separately for the pre-period (2020–2024, divided by 5) and the post-period (2025, divided by 1), yielding two observations per author. Since authors who publish zero papers in a given year are simply absent from the data, a yearly panel would require imputing zeros – but we cannot distinguish true zeros from years before an author began publishing. Averaging over the five-year pre-period and the one-year post-period sidesteps this problem: any author observed in at least one period is assigned zero in the other. See Figure 3 notes for description of comparisons. Only author-level estimates are shown, as paper counts are undefined at the paper level. Thick bars represent 90% confidence intervals; whiskers represent 95% confidence intervals. PPML estimates. Author fixed effects, standard errors clustered by university. Grey = placebo comparisons. Tabular version is in Appendix Table A24.

Table A4: Paper-Level: Gender+Race+Environment Content (Haiku Classification)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0771***						
	(0.0249)						
Post × US High-NSF		-0.106*					
		(0.0640)					
Post × US High-NSF			-0.0741**				
			(0.0365)				
Post × US High-NSF				-0.0685***			
				(0.0239)			
Post × US					-0.00856		
					(0.0306)		
Post × US Low-NSF						0.0142	
						(0.0347)	
Post × UK High-RC							0.0408
							(0.0684)
Observations	10030	1915	2785	14412	11192	9277	1162

Notes: Tabular version of Figure 3 (upper panel). Linear probability model. Robust standard errors in parentheses. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Author-Level: Gender+Race+Environment Content (Haiku Classification, with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0457*						
	(0.0245)						
Post × US High-NSF		-0.0195					
		(0.0350)					
Post × US High-NSF			-0.0290				
			(0.0321)				
Post × US High-NSF				-0.0236			
				(0.0241)			
Post × US					0.00373		
					(0.0246)		
Post × US Low-NSF						-0.0119	
						(0.0383)	
Post × UK High-RC							-0.0360
							(0.0493)
Observations	17422	8040	8621	34706	19749	11468	2278

Notes: Tabular version of Figure 3 (lower panel). Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. See Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Paper-Level: GRE Word Count in Abstract (PPML)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.633***						
	(0.227)						
Post × US High-NSF		-0.822*					
		(0.451)					
Post × US High-NSF			-0.595**				
			(0.299)				
Post × US High-NSF				-0.592***			
				(0.221)			
Post × US					-0.0440		
					(0.227)		
Post × US Low-NSF						0.134	
						(0.265)	
Post × UK High-RC							0.294
							(0.471)
Observations	10030	1915	2785	14412	11192	9277	1162

Notes: Tabular version of Figure A7 (upper panel). PPML estimates. Robust standard errors in parentheses. The dependent variable is the count of words classified into Gender, Race, or Environment categories in the abstract. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Author-Level: GRE Word Count in Abstract (PPML, with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.511***						
	(0.143)						
Post × US High-NSF		-0.196					
		(0.168)					
Post × US High-NSF			-0.230				
			(0.168)				
Post × US High-NSF				-0.282*			
				(0.147)			
Post × US					0.0904		
					(0.145)		
Post × US Low-NSF						0.000356	
						(0.337)	
Post × UK High-RC							-0.137
							(0.405)
Observations	8758	4041	4258	15733	9766	5510	969

Notes: Tabular version of Figure A7 (lower panel). PPML estimates with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is the count of words classified into Gender, Race, or Environment categories in the abstract. See Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Paper-Level: Abstract (binary)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0404						
	(0.0365)						
Post × US High-NSF		-0.107					
		(0.0836)					
Post × US High-NSF			-0.0317				
			(0.0511)				
Post × US High-NSF				-0.0250			
				(0.0354)			
Post × US					0.00418		
					(0.0409)		
Post × US Low-NSF						0.0365	
						(0.0470)	
Post × UK High-RC							0.102
							(0.0887)
Observations	10030	1915	2785	14412	11192	9277	1162

Notes: Tabular version of Figure A8 (upper panel). Linear probability model. Robust standard errors in parentheses. The dependent variable is binary, indicating whether any targeted term appears in the abstract (regardless of context). See Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Author-Level: Abstract (binary, with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0639*						
	(0.0347)						
Post × US High-NSF		-0.0733					
		(0.0553)					
Post × US High-NSF			-0.0544				
			(0.0523)				
Post × US High-NSF				-0.0257			
				(0.0343)			
Post × US					-0.0146		
					(0.0427)		
Post × US Low-NSF						0.0788	
						(0.0642)	
Post × UK High-RC							0.0742
							(0.0765)
Observations	17422	8040	8621	34706	19749	11468	2278

Notes: Tabular version of Figure A8 (lower panel). Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is binary, indicating whether any targeted term appears in the abstract (regardless of context). See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Paper-Level: GRE Word Counts (Qwen3 Full-Text Classification)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.333						
	(0.211)						
Post × US High-NSF		-1.125***					
		(0.417)					
Post × US High-NSF			-0.720**				
			(0.285)				
Post × US High-NSF				-0.363*			
				(0.206)			
Post × US					-0.421*		
					(0.223)		
Post × US Low-NSF						-0.183	
						(0.263)	
Post × UK High-RC							0.574
							(0.451)
Observations	10027	1914	2784	14407	11189	9275	1162

Notes: Tabular version of Figure A9 (upper panel). PPML estimates. Robust standard errors in parentheses. The dependent variable is the count of words classified into Gender, Race, or Environment categories in the full paper text using Qwen3-32B. Page count included as control. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Author-Level: GRE Word Counts (Qwen3 Full-Text, with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.231*						
	(0.118)						
Post × US High-NSF		-0.193					
		(0.204)					
Post × US High-NSF			-0.176				
			(0.193)				
Post × US High-NSF				-0.122			
				(0.110)			
Post × US					-0.0413		
					(0.180)		
Post × US Low-NSF						0.204	
						(0.377)	
Post × UK High-RC							0.127
							(0.441)
Observations	15688	7219	7693	30095	17658	10159	1925

Notes: Tabular version of Figure A9 (lower panel). PPML estimates with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is the count of words classified into Gender, Race, or Environment categories in the full paper text using Qwen3-32B. Page count included as control. See Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A12: Paper-Level: Fulltext counts

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.153						
	(0.102)						
Post × US High-NSF		-0.663***					
		(0.230)					
Post × US High-NSF			-0.285**				
			(0.143)				
Post × US High-NSF				-0.152			
				(0.0991)			
Post × US					-0.145		
					(0.117)		
Post × US Low-NSF						0.0365	
						(0.128)	
Post × UK High-RC							0.531**
							(0.244)
Observations	10027	1914	2784	14407	11189	9275	1162

Notes: Tabular version of Figure A10 (upper panel). PPML estimates. Robust standard errors in parentheses. The dependent variable is the count of targeted words in the full paper text. Page count included as control. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Author-Level: Fulltext counts (with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.103						
	(0.0714)						
Post × US High-NSF		-0.0619					
		(0.0909)					
Post × US High-NSF			-0.0124				
			(0.0989)				
Post × US High-NSF				-0.0198			
				(0.0616)			
Post × US					0.0413		
					(0.0919)		
Post × US Low-NSF						0.299	
						(0.240)	
Post × UK High-RC							0.242
							(0.260)
Observations	17419	8037	8618	34691	19745	11467	2277

Notes: Tabular version of Figure A10 (lower panel). PPML estimates with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is the count of targeted words in the full paper text. Page count included as control. See Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Paper-Level: GRE Content (2024 only pre-period)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0939***						
	(0.0340)						
Post × US High-NSF		-0.0834					
		(0.0881)					
Post × US High-NSF			-0.0858*				
			(0.0483)				
Post × US High-NSF				-0.0840**			
				(0.0329)			
Post × US					-0.00457		
					(0.0391)		
Post × US Low-NSF						0.0101	
						(0.0435)	
Post × UK High-RC							-0.0111
							(0.0933)
Observations	3298	608	905	4870	3687	3079	389

Notes: Tabular version of Figure A11. Linear probability model. Robust standard errors in parentheses. Pre-treatment period restricted to 2024 only. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A15: Paper-Level: Treatment Effects by Content Category (US High-NSF vs US Low-NSF)

	(1) GRE	(2) Gender	(3) Race	(4) Environment	(5) Econ Ineq	(6) Other
Post × US High-NSF	-0.0771***	-0.0290	-0.0262*	-0.0167	0.0149	-0.00249
	(0.0249)	(0.0182)	(0.0150)	(0.0151)	(0.0331)	(0.0168)
Observations	10030	10030	10030	10030	10030	10030

Notes: Tabular version of Figure 4 (upper panel). Linear probability model. Robust standard errors in parentheses. US High-NSF vs US Low-NSF only. See Figure 4 notes for description of categories.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A16: Author-Level: Treatment Effects by Content Category (US High-NSF vs US Low-NSF, with Author FE)

	(1) GRE	(2) Gender	(3) Race	(4) Environment	(5) Econ Ineq	(6) Other
Post × US High-NSF	-0.0457*	-0.0158	-0.0242**	-0.00591	-0.0105	-0.00355
	(0.0245)	(0.0160)	(0.0108)	(0.0120)	(0.0226)	(0.0136)
Observations	17422	17422	17422	17422	17422	17422

Notes: Tabular version of Figure 4 (lower panel). Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. US High-NSF vs US Low-NSF only. See Figure 4 notes for description of categories.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A17: Paper-Level: NSF/SES Funding Exclusion (GRE Topic)

	(1) Full sample	(2) Excl NSF/SES	(3) Excl NSF	(4) Excl SES
Post × US High-NSF	-0.0713**	-0.0735**	-0.0663**	-0.0749***
	(0.0290)	(0.0294)	(0.0296)	(0.0290)
Observations	7503	7012	7168	7259

Notes: Tabular version of Figure 5 (upper panel). Linear probability model. Robust standard errors in parentheses. US High-NSF vs US Low-NSF only. See Figure 5 notes for description of exclusion criteria.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A18: Author-Level: NSF/SES Funding Exclusion (GRE Topic, with Author FE)

	(1) Full sample	(2) Excl NSF/SES	(3) Excl NSF	(4) Excl SES
Post $\times$ US High-NSF	-0.0483*	-0.0378	-0.0373	-0.0458*
	(0.0263)	(0.0267)	(0.0270)	(0.0257)
Observations	13343	12164	12561	12760

Notes: Tabular version of Figure 5 (lower panel). Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. US High-NSF vs US Low-NSF only. See Figure 5 notes for description of exclusion criteria.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A19: Heterogeneity: Author-Level GRE Content by Subsample

	(1) All US	(2) Senior	(3) Junior	(4) White	(5) Minority	(6) Not in Census	(7) Male	(8) Female
Post × High-NSF	-0.0458*	-0.0593*	-0.0363	-0.0622*	-0.00201	-0.0459	-0.0388	-0.0960*
	(0.0245)	(0.0308)	(0.0343)	(0.0338)	(0.0445)	(0.0344)	(0.0253)	(0.0510)
Observations	17403	8967	7798	8946	3612	4845	12625	3561

Notes: Tabular version of Figure 6. Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. US High-NSF vs US Low-NSF only. See Figure 6 notes for description of subsamples.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A20: Paper-Level: GRE Content with Red/Blue State × Post Control

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0748***						
	(0.0248)						
Post × US High-NSF		-0.130**					
		(0.0639)					
Post × US High-NSF			-0.0983***				
			(0.0364)				
Post × US High-NSF				-0.0700***			
				(0.0236)			
Post × US					-0.0190		
					(0.0308)		
Post × US Low-NSF						0.00736	
						(0.0350)	
Post × UK High-RC							0.0408
							(0.0684)
Observations	10030	1915	2785	14412	11192	9277	1162

Notes: Tabular version of Figure A12 (upper panel). Linear probability model. Robust standard errors in parentheses. Controls for Red State × Post interaction. See Figure A12 notes for details and Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A21: Author-Level: GRE Content with Red/Blue State × Post Control

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0427*						
	(0.0244)						
Post × US High-NSF		-0.0292					
		(0.0386)					
Post × US High-NSF			-0.0389				
			(0.0361)				
Post × US High-NSF				-0.0257			
				(0.0245)			
Post × US					-0.00397		
					(0.0252)		
Post × US Low-NSF						-0.0173	
						(0.0383)	
Post × UK High-RC							-0.0360
							(0.0493)
Observations	17422	8040	8621	34706	19749	11468	2278

Notes: Tabular version of Figure A12 (lower panel). Linear probability model with author fixed effects. Standard errors clustered by university in parentheses. Controls for Red State × Post interaction. See Figure A12 notes for details and Figure 3 notes for description of comparisons.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A22: Paper-Level: Paper Length (Number of Pages)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0118						
	(0.0298)						
Post × US High-NSF		0.0321					
		(0.0672)					
Post × US High-NSF			-0.0395				
			(0.0405)				
Post × US High-NSF				-0.00308			
				(0.0289)			
Post × US					-0.0282		
					(0.0321)		
Post × US Low-NSF						-0.0505	
						(0.0361)	
Post × UK High-RC							-0.0964
							(0.0702)
Observations	10027	1914	2784	14407	11189	9275	1162

Notes: Tabular version of Figure A13 (upper panel). PPML estimates. Robust standard errors in parentheses. The dependent variable is the number of pages. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A23: Author-Level: Paper Length (Number of Pages, with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0178						
	(0.0254)						
Post × US High-NSF		-0.0907*					
		(0.0536)					
Post × US High-NSF			-0.0748				
			(0.0456)				
Post × US High-NSF				-0.0312			
				(0.0232)			
Post × US					-0.0570		
					(0.0400)		
Post × US Low-NSF						0.00438	
						(0.0250)	
Post × UK High-RC							0.0766
							(0.0531)
Observations	17419	8037	8618	34691	19745	11467	2277

Notes: Tabular version of Figure A13 (lower panel). PPML estimates with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is the number of pages. See Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A24: Author-Level: Avg. Papers per Year (with Author FE)

	(1) vs US Low	(2) vs UK High	(3) vs UK All	(4) vs All Other	(5) US vs UK	(6) Low vs Low	(7) UK H vs L
Post × US High-NSF	-0.0670						
	(0.0686)						
Post × US High-NSF		-0.0762					
		(0.109)					
Post × US High-NSF			-0.142				
			(0.0941)				
Post × US High-NSF				-0.0930			
				(0.0602)			
Post × US					-0.0368		
					(0.0884)		
Post × US Low-NSF						-0.122	
						(0.0840)	
Post × UK High-RC							-0.126
							(0.124)
Observations	12938	6274	6914	33728	15088	9028	2102

Notes: Tabular version of Figure A14. PPML estimates with author fixed effects. Standard errors clustered by university in parentheses. The dependent variable is the average number of papers per author per year (pre: 2020–2024, post: 2025; two observations per author). See Figure A14 notes for details and Figure 3 notes for description of comparisons.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$