Check for updates



SOCIAL SCIENCES

# Ideological bias in the production of research findings

George J. Borjas[1,2]* and Nate Breznau[3]

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

**When studying policy-relevant topics, researchers' policy preferences may shape analytical decisions and results interpretations. Detecting this bias is challenging because the research process is not normally part of an observed experimental setting. Our study exploits an opportunity to observe 158 researchers working independently in 71 teams during an experiment. After being asked their position on immigration policy, they used the same data to answer the same empirical question: Does immigration affect public support for social welfare programs? The researchers estimated 1253 alternative regression models, and the estimated impacts ranged from strongly negative to strongly positive. We find that teams composed of pro-immigration researchers estimated more positive impacts of immigration on public support for social programs, while anti-immigration teams estimated more negative impacts. The differences arise because different teams adopted different model specifications. The underlying research design decisions are the mechanism through which ideology enters the process of producing parameter estimates.**

## INTRODUCTION

Consumers of empirical research in the social and behavioral sciences are accustomed to variation in estimates of important theoretical and policy-relevant parameters. For example, existing estimates measuring the impact of a rise in the minimum wage on the employment rate of affected workers (*1*, *2*) or the impact of immigration on native worker opportunities (*3*–*5*) range from strongly negative to strongly positive. Metascience offers insights in how the research process generates these dispersions in findings (*6*). There is evidence of confirmation bias, where researchers attempt to hack their results to find what they wanted in the first place (*7*–*9*); publication bias, where editors and reviewers filter out findings for reasons that are not purely scientific (*10*, *11*); the "file drawer problem," where some research findings never get submitted for publication (*12*); the use of questionable data and faking of results (*13*); and the noise and errors that enter the research production function because of human judgment and fallibility (*14*). In this study, we focus on how the political ideology of scientists may bias their policy-related research results and interpretations (*15*–*17*).

As the production of research findings does not take place in an observed experimental setting, it is difficult to isolate the potential role of ideological bias. The problem is further confounded because ideological bias can enter the research process in many ways at different stages, including the framing of the hypothesis and the design of the research methodology.

Despite its potential importance, few studies directly examine the role of ideology in the research process. Instead, some experiments document ideological bias in peer review of the research process. In the classic study by Abramowitz *et al.* (*18*), a sample of psychologists reviewed one of two randomly assigned manuscripts that were otherwise identical except for the conclusion. In particular, the relative well-being of college student activists who occupied a campus building was either higher or lower than a nonactivist sample of the same student population. Left-leaning psychologists gave a significantly higher ranking to manuscripts where activists reported higher levels of well-being. Similarly, Finseraas *et al.* (*19*) conducted an experiment where a sample of Norwegian researchers evaluated a fictitious study on majority-minority relations. Although all reviewers evaluated the same research design, the conclusion of the study had been randomized to either support "left-liberal" implications or not. The researchers who evaluated the version with the left-liberal conclusion ranked that research design as being of higher quality. Finally, one study examined ideology more directly in the research production process via a textual analysis of published studies in economics (*17*). It found a correlation between the language used in the articles and the political partisanship of the researchers.

We document the role of ideology in the production of research findings by taking advantage of the unique crowdsourced experiment designed by N. Breznau, E. Mark Rinke, and A. Wuttke (*20*) (henceforth BRW). In their experiment, 71 research teams (comprising 158 researchers) received the same data to test the specific hypothesis that immigration reduces support for social policies among the public. The teams had no control over the data to be used or the hypothesis to be tested. Past research on this hypothesis is interdisciplinary and vast. Collectively, it suggests that the correlation between immigration and support for social programs may be negative (*21*–*25*), may be zero (*26*, *27*), or may be positive (*28*–*30*). As part of the experiment and before any empirical analysis was done, the participating researchers were asked about their attitudes toward immigration; specifically, whether immigration laws "should be made tougher" or "should be relaxed."

The research teams produced wildly different estimates of the impact of immigration. As BRW [page 5 of (*20*)] note, "no two teams arrived at the same set of numerical results or took the same major decisions during data analysis." Our study examines this variation to determine whether the dispersion depends on researcher priors about immigration. Our analysis documents that such a correlation exists, a correlation overlooked by BRW in their analysis of the data for substantively interesting and important reasons (discussed shortly). Research teams that had strong pro-immigration sentiments were more likely to obtain positive parameter estimates, suggesting that immigration increases public support for social programs, which social and behavioral scientists consider to be an indicator of increased social cohesion (*31*). Conversely, strongly anti-immigration teams were more likely to obtain negative estimates, potentially suggesting that immigration reduces social cohesion.

[1]Harvard Kennedy School, 79 John F. Kennedy Street, Cambridge, MA 02138, USA. [2]National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, USA. [3]German Institute for Adult Education–Leibniz Centre for Lifelong Learning, Department of Organization and Program Planning, Heinemannstraße 12-14, 53175 Bonn, Germany.
*Corresponding author. Email: george_borjas@hks.harvard.edu

Downloaded from https://www.science.org on January 01, 2026

We demonstrate that the combinations of five specific research design decisions account for about two-thirds of the difference in the impacts reported by pro- and anti-immigration teams. The research design choices that could uniquely produce large negative effects were taken almost exclusively by anti-immigration teams and vice versa for pro-immigration teams. In short, research design is endogenous and provides the mechanism through which ideology enters the process of producing research findings. Last, dispersion in the quality of the model specifications adopted by different teams, as measured by a random and double-blind review of each modeling strategy by other researchers in the experiment, shows that anti- and pro-immigration teams received lower peer referee scores than moderate teams.

As noted above, our finding that ideology matters conflicts with the BRW [page 5 of (*20*)] conclusion that "researcher characteristics do not explain outcome variance." The reason for the difference is substantively important: The preferred statistical analysis in BRW was a regression model [(*20*), BRW's Supplemental Appendix, and table S5] where the dependent variable—the team's estimate of the impact of immigration on social cohesion—was regressed on a variable measuring the team's ideology jointly with a vector of many other variables that described characteristics of the model specification used by the team [e.g., logit or ordinary least squares (OLS), inclusion of country and year fixed effects, etc.]. The ideology variable was not statistically significant in this expanded specification. Our analysis, however, documents that this regression model is misspecified if its purpose is to determine whether ideology plays an independent role in the production of research findings. The variables indicating various aspects of the regression specification are endogenous and are the mechanism through which ideology influences statistical estimates. Different researchers choose different specifications, and this choice is influenced by ideology. In hindsight, the "kitchen sink" regression specification of BRW, while useful in measuring how much of the variance in the estimated impact of immigration is explained by "researcher characteristics," cannot identify the impact of ideology on the estimated parameters.

In summary, our exploratory analysis documents that ideology has a statistically significant effect on the production of research findings that is robust across model specifications, that much of this effect works through the decision-making process that leads researchers to choose different modeling strategies for examining the same data and answering the same question, and that the identification of the ideology effect requires a careful statistical analysis where the potential endogeneity of the research process is explicitly considered.

## RESULTS

### Baseline evidence

The raw data suggests that pro-immigration teams tend to adopt research strategies that lead to the conclusion that immigration increases social cohesion, while the opposite is true for anti-immigration teams. The average marginal effect (AME) density distributions in Fig. 1C illustrate the pattern. There is much more mass in the negative (positive)

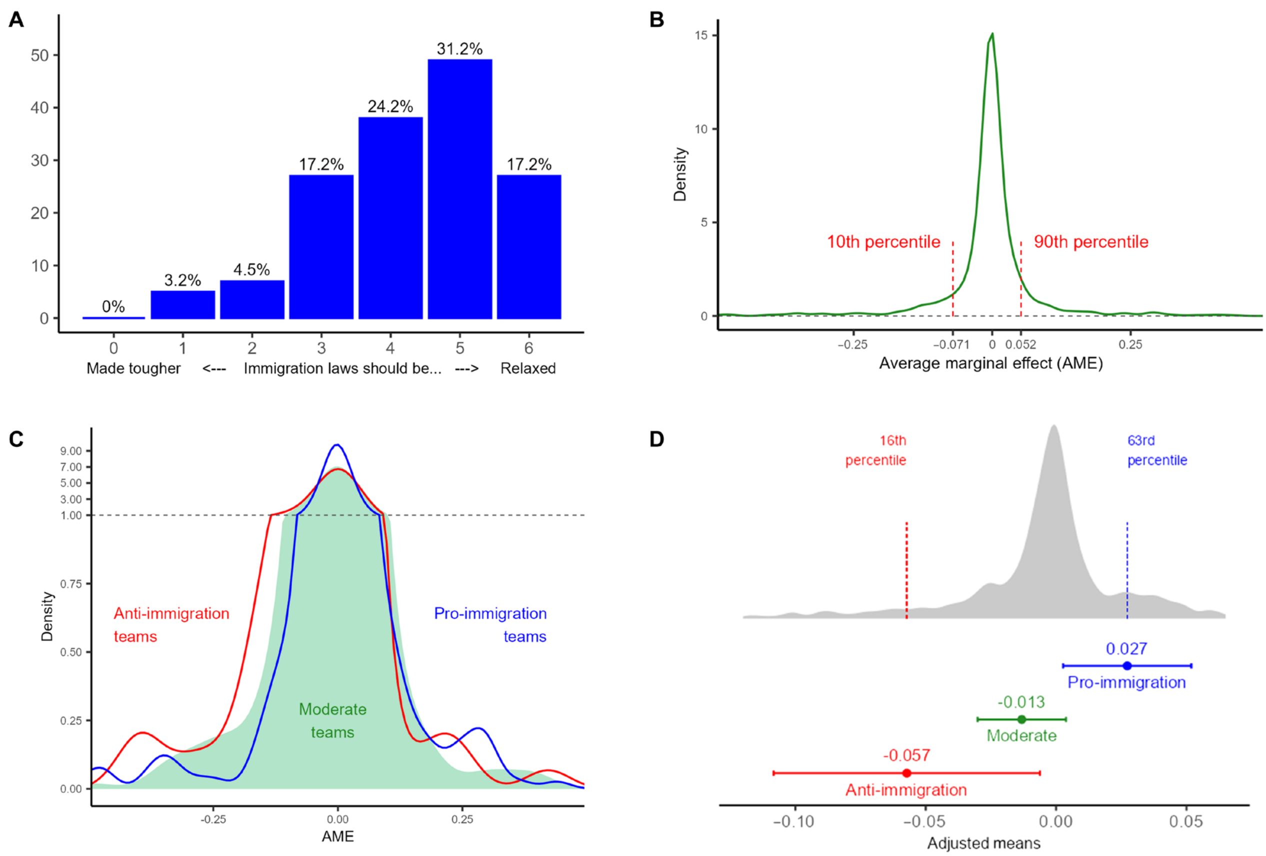


**Fig. 1. Ideology and policy-related estimates from experiment involving 71 teams.** The experiment of BRW measured ideological policy positioning of 158 researchers in 71 teams (**A**) who then tested the hypothesis that immigration reduces support for social policy using the same data leading to 1253 estimates of the AME (**B**). There is a strong distributional association between ideology and AME (**C**). This pattern remains after weighting by the number of models per team and adjusting for covariates (**D**).

Downloaded from https://www.science.org on January 01, 2026

tail for the anti-immigration (pro-immigration) teams. The summary statistics reported in table S1 further document differences in the estimated AME across team types. The raw differences in the mean AME are small: It is slightly positive (0.014) for pro-immigration teams, slightly negative (−0.008) for moderate teams, and most negative (−0.019) for the anti-immigration teams. Differences are starker if we focus on the tails of the distributions and significance of the estimates. Among the AME estimates produced by pro-immigration teams, for example, 5.9% are positive and significant at the 5% level in a one-tailed test (i.e., $|t| > 1.645$), and 2.8% are negative and significant. In contrast, among anti-immigration teams, 3.7% are positive and significant, and 11.9% are negative and significant.

Note that regardless of the team's ideology, the AME frequency distributions tend to cluster around zero. As noted earlier, the starting task in the empirical analysis conducted by all teams was to reproduce the Brady-Finigan (*27*) null finding. It would not be surprising if the distributions are influenced by the results from that initial task. Some of the deviation we observe from the null finding arises because anti-immigration teams seem to have made the effort to extend that initial analysis in ways that produced negative estimates of the AME, while the opposite is true for the pro-immigration teams.

The direct link between ideology and the AME can be documented by estimating regressions that relate the estimated AME to a vector of team-specific variables. The regression model is

$$\text{AME}_{rs} = \alpha I_r + \text{controls} + \text{error} \tag{1}$$

where $\text{AME}_{rs}$ is the estimate obtained by team $r$ using regression specification $s$ and $I_r$ gives a measure of team $r$'s ideology toward immigration. The regressions are weighted by the inverse of the number of models estimated by the team, and standard errors are clustered at the team level.

The regressions include controls for team size and researcher experience using relevant empirical methods and conducting research in the topical area (with detailed descriptions of all variables given in the Supplementary Materials). The regressions also include controls for the (sometimes mixed) disciplinary background of the team members (e.g., sociology, political science, economics, etc.). This variable is crucial because the educational trajectory of a researcher may have a causal association with their political ideology (and/or the processes that develop this ideology) and expressed policy preferences. It is also evident that different discipline and degree programs expose researchers to different approaches to model specification and data analysis. Disciplinary background, therefore, may be a confounder variable that should be included in any correctly specified model to avoid omitted variable bias.

Most researchers are either sociologists (55.4%) or political scientists (27.4%). However, 57 of the 71 teams have more than one researcher, and 36 of those 57 combine myriad disciplines. Our baseline specification includes fixed effects indicating the discipline of the lead author, fixed effects for two-person teams with researchers from different disciplines, and fixed effects for three-person teams indicating if the majority are sociologists or political scientists and if the lead author's discipline differs from that of the others

Columns 1 to 3 in Table 1 report key regression coefficients using three alternative specifications for the ideology variable. Column 1 shows that going from one extreme of the mean sentiment index to the other (from a "1" to a "6" in the scale) increases the estimated AME by 0.054 points (with a standard error of 0.029). Column 2

Downloaded from https://www.science.org on January 01, 2026

**Table 1. Impact of ideology on the estimated AME.** Note that $^{*}P < 0.1$ and $^{**}P < 0.05$. Full regression results are in tables S2 to S4. Standard errors reported in parentheses and are clustered at the team level. Columns 1 to 3 use AME as the dependent variable, columns 4 to 6 use extreme negative and significant effects (<10th percentile and $P < 0.05$) and linear probability models (LPMs), and columns 7 to 9 use extreme positive and significant effects (>90th percentile and $P < 0.05$) and LPMs. For alternative ideology measures, we use the mean score for each team on the ideology question (columns 1, 4, and 7), the percent of each team that are anti- or pro-immigration (columns 2, 5, and 9), and the categorization of teams as pro-immigration, moderate, or anti-immigration (columns 3, 6, and 9). All regressions have 1253 observations. The regressors also include the team's index of statistical skills, topical experience, team size fixed effects, and the discipline composition of the team members. The unit of observation in the regressions is a model estimated by a particular team. We weigh all regressions by the inverse of the number of models reported by the team. If all covariates are constant within a team, then the regression coefficients would be numerically identical to those obtained when the regression is estimated using team-level data instead. $R^2$, coefficient of determination.

| | Dependent variable | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | AME | | | Extreme negative and significant AME | | | Extreme positive and significant AME | | |
| **Variable:** | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** | **(9)** |
| Mean immigration sentiment | 0.011*<br>(0.006) | – | – | −0.051**<br>(0.021) | – | – | 0.019*<br>(0.012) | – | – |
| % of team that is anti-immigration | – | −0.050**<br>(0.025) | – | – | 0.081<br>(0.103) | – | – | −0.130*<br>(0.057) | – |
| % of team that is pro-immigration | – | 0.023<br>(0.019) | – | – | −0.161**<br>(0.064) | – | – | −0.016<br>(0.041) | – |
| Anti-immigration team | – | – | −0.057*<br>(0.031) | – | – | 0.150<br>(0.123) | – | – | −0.070*<br>(0.037) |
| Pro-immigration team | – | – | 0.027*<br>(0.015) | – | – | −0.124**<br>(0.044) | – | – | 0.017<br>(0.034) |
| Difference: Pro − Anti | – | 0.074**<br>(0.024) | 0.085**<br>(0.031) | – | −0.242*<br>(0.107) | −0.274*<br>(0.125) | – | 0.114*<br>(0.045) | 0.087*<br>(0.035) |
| $R^2$ | 0.062 | 0.065 | 0.071 | 0.096 | 0.115 | 0.133 | 0.087 | 0.090 | 0.089 |

relates the AME to the percentage of the team members that are either pro- or anti-immigration. The larger the representation of anti-immigration (pro-immigration) researchers in the team is, the lower (higher) the estimated AME is. The difference in the estimated AME between teams composed exclusively of either pro- or anti-immigration researchers is 0.074 (0.024) points. Last, column 3 estimates the difference in the adjusted AME across the three types of teams defined above. The AME estimated by pro-immigration teams is 0.027 (0.015) points higher than the estimate of the moderate teams. In contrast, the AME estimated by anti-immigration teams is −0.057 points lower. The difference between the pro- and anti-immigration teams is 0.085 (0.031) points and statistically significant (with a $P$ value of 0.008). The impact of ideology on parameter estimates is not only statistically significant but also numerically important. Figure 1D shows that the difference in the estimated AME between an anti-immigration team and a pro-immigration team is equivalent to moving from the 16th percentile of the (adjusted) AME distribution to the 63rd percentile (with point estimates −0.057 and 0.027, respectively). This shift represents 68% of the distance between the 10th and 90th percentiles shown in Fig. 1B.

The remaining columns of Table 1 report the coefficients from linear probability models where the dependent variable is a binary indicator set to unity if the estimated AME is in one of the tails (either below the 10th or above the 90th percentile) and statistically significant at $P < 0.05$ (and zero otherwise). The statistical impact of ideology is notable. Anti-immigration teams are less likely to estimate significant effects in the positive tail, while pro-immigration teams are less likely to estimate significant effects in the negative tail. The probability that an anti-immigration team estimates a sizable negative AME is 27.4 (12.5) percentage points higher than that of a pro-immigration team. Similarly, the probability that a pro-immigration team estimates a sizable positive AME is 8.7 (3.5) percentage points higher than that of an anti-immigration team (columns 8 and 12 in Table 1). These results reinforce the insight that given the starting point of replicating the Brady-Finnigan null result, it is the most ideological teams that tend to depart from this finding and adopt specifications that produce an AME estimate in either tail.

The Supplementary Materials present additional regressions documenting that our results are robust to various specifications. Tables S2 to S4 show that the baseline results in Table 1 are unchanged if the regressions also include a variable measuring prior belief in the validity of the hypothesis being tested (panels A and B) or if the regressions are weighted by "referee scores" from a blind peer review of each model specification (panel B), a process discussed in more detail below. Table S5 reestimates the regressions using a dependent variable that simply indicates if the estimated AME is positive or negative (excluding the models that estimated an AME numerically very close to zero, i.e., between −0.015 and +0.015). This table confirms our substantive conclusion: Positive (negative) coefficients are more likely to be estimated by pro-immigration (anti-immigration) teams.

Tables S6 and S7 examine the sensitivity of the evidence to controlling for the role of disciplinary background. Table S6 shows that the baseline results are similar when we use two alternative strategies for summarizing the team's disciplinary background (e.g., using only the discipline of the corresponding author or fixed effects that capture all discipline combinations among the up-to-three researchers per team). Table S7 documents step-wise robustness when adding the control variables and illustrates the potential role of omitted variable bias in specifications that exclude disciplinary background. Recall that the summary statistics indicated that the unadjusted differences in the mean AME across the three types of teams were relatively small. Nevertheless, the regressions show the differences become large and significant after controlling for the team's disciplinary composition, statistical skill, and topic experience (as well as team size). The key control that produces the large and significant ideology impact is the vector of discipline fixed effects. As table S7 shows, the baseline regressions reported in Table 1 would be almost identical if the only other regressors were those fixed effects. As noted earlier, this vector partly accounts for the fact that the various disciplines differ in their methodological training and in their preferred methods, expectations, and evaluation systems (*32*). The large differences may also be related to the (unobserved) existence and formation of different ideological viewpoints in different disciplines. The disciplinary fixed effects then help isolate the ideology effect by comparing the research findings produced by researchers who have different immigration sentiments after accounting for the potentially confounding interactions between disciplinary training and ideological perspectives.

The last three tables in the Supplementary Materials examine if the evidence is sensitive to how the team's ideology is defined. Table S8 reports the difference in the estimates between pro-immigration teams and all other teams, bypassing the need to separate out the small number of teams with anti-immigration sentiments. It is evident that pro-immigration teams are more likely to estimate positive AMEs than all other teams. To further probe into the robustness of the ideology effect, table S9 relates the estimated AME to the immigration sentiments of the most senior member of the team (defined as the researcher with the highest level of statistical skills). This exercise allows for the possibility that senior researchers have a disproportionate influence in determining the nature of the team's research product. The data again reveal a strong ideological effect, with a significant difference in estimated AMEs between teams "led" by persons with anti-immigration and pro-immigration preferences. Table S10 shows that the substantive results are unchanged if the data are restructured so that the unit of observation is a researcher-model dyad (rather than the team-model level). This restructuring allows the analysis to directly relate the estimated AME to a single researcher's ideology or discipline. The regressions show that individual researchers who are at the anti- or pro-immigration extremes of the immigration sentiment distribution were more likely to be in teams that reported negative or positive effects of immigration, respectively.

The regression models used in our analysis represent a cross section of the logically justifiable models that would be normally used in this type of exercise in empirical social science. As an agnostic robustness check, we estimated the ideology effect in the 883 models that represent all combinations of the regression specifications used in Table 1 and tables S2 to S10. We find that ideology has a statistically significant effect (with $P < 0.10$) in 88.2% of the models (with the fraction increasing to 92.4% if we exclude the models that may suffer from omitted variable bias because the disciplinary fixed effects are not in the regression). Figure S3 shows the specification curve. The aggregated evidence, therefore, seemingly satisfies a standard of robustness that follows a Bonferroni logic: In the presence of a true effect, a handful of models (~10%) will not reveal it.

## Endogeneity of research design

The regression analysis implies that teams at the extremes of the immigration sentiment distribution use different regression specifications. It

Downloaded from https://www.science.org on January 01, 2026

is important to emphasize that the regressions estimated in Table 1 do not include any variables that describe the actual specification of the regression model used to produce a particular estimate of the AME. After "playing around" with the data, a researcher might infer how particular variables and estimation techniques influence the sign and magnitude of the estimated AME. For example, the International Social Survey Program (ISSP) asked several questions about the government's responsibility to provide specific services, such as housing or health programs. The results could be sensitive to which subset of programs is used as the dependent variable. Similarly, the immigrant supply shock can be measured as a share of the population that is foreign-born or as net migration per year. The choice of variables and estimation techniques is one mechanism through which ideological bias might influence the estimate of the AME. Put simply, research design is endogenous.

The BRW experimental data record the outcome of 103 different specification decisions taken by more than one team, such as the exact definition of the dependent variable; the measure of the immigrant shock; the set of European countries and the waves of the ISSP used in the analysis; and the choice of statistical methods such as linear probability models, multinomial logit, random effects, etc. It turns out that unique combinations of decisions made along five dimensions produce much of the variation in the mean AME estimated by the three types of teams. These key decisions are:

1. The ISSP records public attitudes toward the government provision of various types of programs (jobs, health, etc.). Are these responses aggregated to form a single dependent variable (e.g., by averaging or estimating a factor index)?
2. Is immigration measured as a stock or a flow?
3. Does the analysis use multilevel modeling to account for variation in country-year units?
4. Do the regressions use data for all countries in the ISSP?
5. Does the analysis use the 2016 wave of the ISSP (in addition to the 1996 and 2006 waves used in the original Brady-Finnigan study)?

The combination of these decisions produces 58 alternative regression specifications with nonempty cells (see table S11 for summary statistics). For each of these specifications, we calculated the "expected AME," defined as the mean AME in the subset of models using that specification. Figure 2A ranks the 58 specifications from lowest to highest expected AME and shows the frequency of adopting each specification. Notably, anti-immigration teams were the only ones that used specifications producing the lowest expected AMEs and that pro-immigration teams were the only ones using specifications that produced the highest. Figure 2B illustrates the frequency distribution of the expected AME for each type of team. The density function for the anti-immigration teams has little mass for expected AMEs above 0.0, while the density function for pro-immigration teams has little mass below −0.1.

We reestimated the basic regression models to determine how the team's immigration ideology affects the expected AME distribution. The coefficients reported in Table 2 are roughly similar to the analogous coefficients reported in columns 1 to 3 of Table 1 (from regressions that use the actual AME). Column 3 of Table 2 shows that anti-immigration teams choose specifications that, on average, have an expected AME that is −0.044 (0.024) points lower than the typical model estimated by a moderate team, while pro-immigration teams choose specifications that have an AME that is 0.014 (0.008) higher. The difference in the AME estimated by the two extreme types of teams is 0.058 (0.025) points ($P = 0.025$).

The rough similarity between Tables 1 and 2 suggests that inter-team differences in research design decisions along five dimensions account for much of the (mean) AME difference between team types. The regression in column 3 of Table 1 implies that the difference in actual AMEs between the two extreme types of teams is 0.085, while the corresponding regression in column 3 of Table 2 implies that the difference in expected AME is 0.058. Hence, the five design choices account for 68% (or 0.058 ÷ 0.085) of this difference. The analogous calculation using the coefficients from the regressions reported in columns 2 of Tables 1 and 2 that use the fraction of the team's members that are pro- or anti-immigration implies that those choices account for 54% of the difference.

In summary, the endogenous research design choices made by teams with immigration ideologies in either tail of the immigration sentiment distribution happen to produce parameter estimates that seem consistent with the teams' pre-existing ideological priors. Note that we are careful to say "choices" as plural here because a key finding of the original BRW study was that no single specification choice mattered much (e.g., logit or OLS). A more recent study, which simulates the multiverse of all possible choices in the BRW experimental data, confirms that indeed no single specification choice explains much of the variation, but that higher-order interactions of those choices explain much of the meaningful variation in outcomes (*32*). Put differently, research design decisions that influence the results must be the result of combinations of specification decisions, not single decisions. Our analysis of these combinations thus provides a straightforward depiction of the mechanism through which ideology influences research outcomes.

As noted earlier, our finding that ideology affects the estimate of the AME (because ideology influences the team's research design) conflicts with the original BRW conclusion [page 5 of (*20*)] because our empirical analysis explicitly accounts for the possibility that variables indicating research design choices are endogenous to the process. The examination of any data by any researcher will likely reveal certain patterns that become accentuated or less visible as the researcher explores alternative statistical models. We have shown that the final specification decisions made by researchers are correlated with pre-existing priors about the topic being examined.

Although it would be of interest to conduct a study of exactly how researchers end up using a specific "preferred" specification, the experimental data do not allow examination of this crucial question. For example, we do not know if researchers "consciously or unconsciously" gravitate to models that support their preferred conclusions. Although the study of the production of research findings is in its infancy, our study raises crucial questions that are important for understanding the partly subjective nature of empirical research.

Last, each team's research design was reviewed anonymously by four or five other researchers participating in the experiment, and BRW constructed a "referee score" for each model by averaging these reviews. These data allow us to examine the possibility that immigration ideology not only biases the estimated AME, but that some of that bias arises because ideology affects research quality. If this conjecture is correct, then the impact of immigration ideology on research quality may not be monotonic but might instead show up in both tails of the immigration sentiment distribution.

Figure 3 illustrates the distribution of standardized referee scores for each of the three types of teams. It shows that both the anti- and pro-immigration teams have raw distributions of referee scores that

Downloaded from https://www.science.org on January 01, 2026

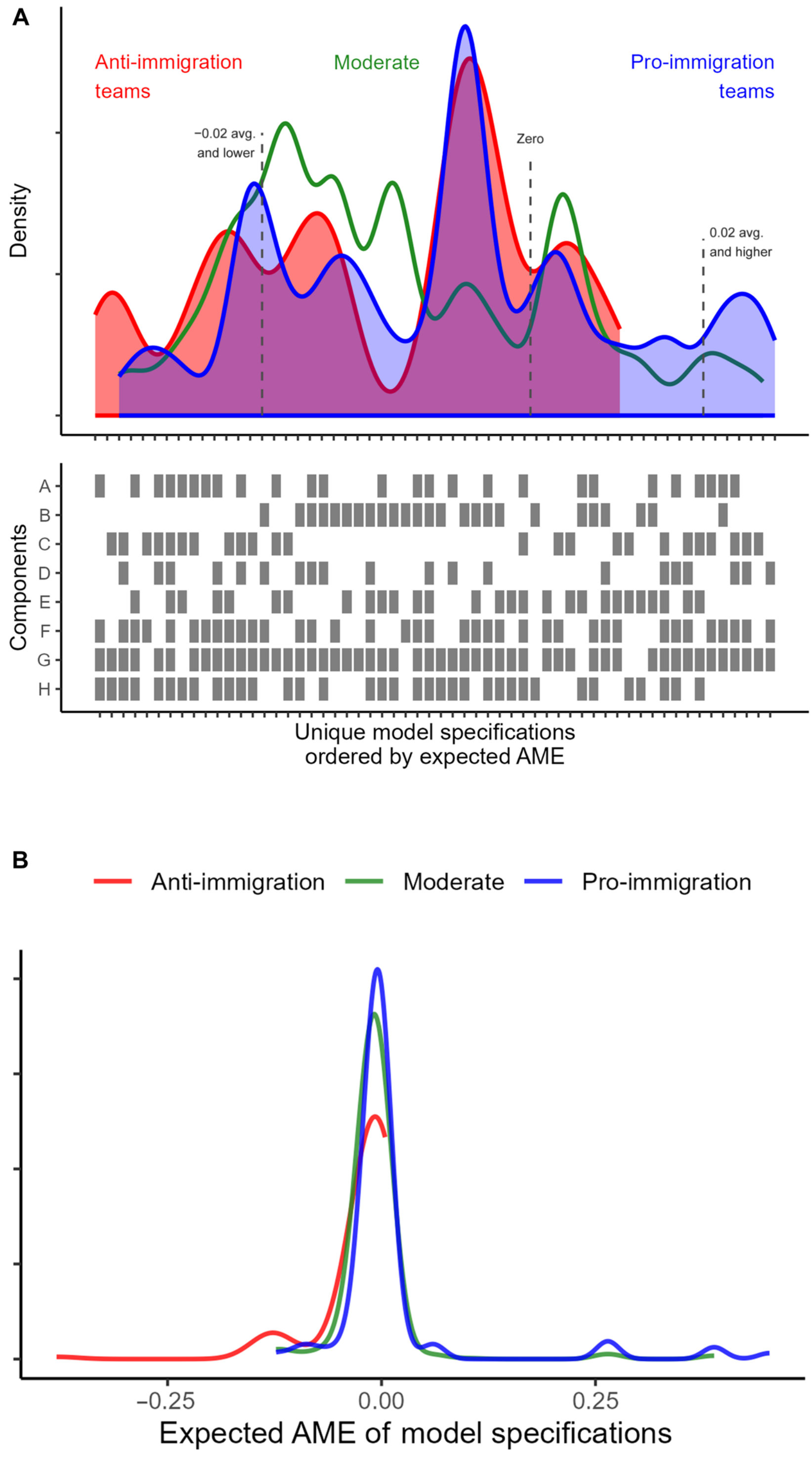


**Fig. 2. Expected outcomes of AME based on five central modeling decisions.** (**A**) distribution of model specifications by team ideology ranked by order of expected AME, calculated as the mean AME for these models. The decisions are: (i) scaling: A = composite dependent variable; (ii) test variable: B = immigration measured as stock (% foreign-born), C = immigration measured as flow (net migration); (iii) model structure: D = multilevel modeling for variation at country-year level; (iv) countries: E = all available countries in the data included; and (v) data waves: F = data from the 1996 wave included, G = data from the 2006 wave included, and H = data from the 2016 wave included. (**B**) distribution with unique specifications plotted by expected AME.

Downloaded from https://www.science.org on January 01, 2026

**Table 2. Determinants of expected AME.** Note that $*P < 0.10$ and $**P < 0.05$. Standard errors in parentheses are clustered by number of models per team. Regressions weighted by the inverse of the number of models reported by the team. All regressions have 1253 observations. The regressors also include the team's index of statistical skills, topical experience, team size fixed effects, and the discipline composition of the team members. The unit of observation in the regressions is a model estimated by a particular team. See table S10 for full regressions and robustness checks.

| | Specification | | |
|---|---|---|---|
| **Variable:** | **(1)** | **(2)** | **(3)** |
| Mean immigration sentiment | 0.007* (0.004) | – | – |
| % of team that is anti-immigration | – | −0.026 (0.019) | – |
| % of team that is pro-immigration | – | 0.014 (0.014) | – |
| Anti-immigration team | – | – | −0.044* (0.024) |
| Pro-immigration team | – | – | 0.014* (0.008) |
| Difference: Pro − Anti | – | 0.040* (0.020) | 0.058** (0.025) |
| $R^2$ | 0.104 | 0.110 | 0.140 |

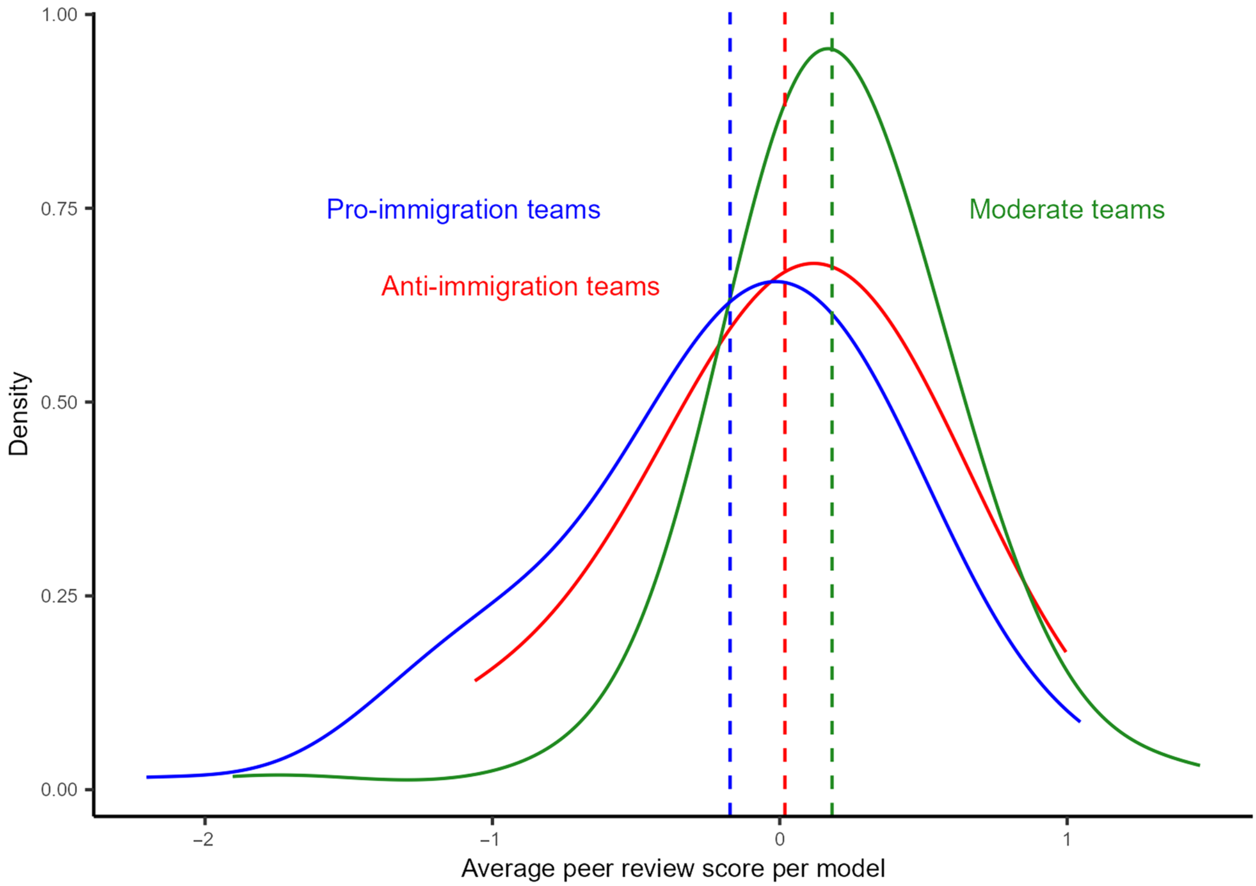


**Fig. 3. Distribution of standardized referee score by type of team.** The figures present the raw distribution of standardized peer review scores for 1215 models across 71 teams by team ideology; the vertical dotted lines represent the raw means for each type of team.

Downloaded from https://www.science.org on January 01, 2026

lie to the left of the distribution of the moderate teams. The moderate teams obtain the highest mean score of 0.35, as compared to 0.03 for the anti-immigration teams and −0.33 for the pro-immigration teams. Table S12 shows that these differences in referee scores across the various types of teams remain even after controlling for the set of regressors introduced earlier (i.e., statistical skills, topic experience, and discipline fixed effects). Note that the documented differences in research quality across the three types of teams are not affected by potential ideological bias on the part of referees. The reviewing process was double-blinded and random, so that even if referees preferred specifications that would confirm their own biases, the average referee score achieved by any model would not be affected. Further, the reviewers did not have access to the actual results of other teams, only to a verbal description of the statistical model and variables to be included. The evidence thus suggests that peers in the experiment were aware that "nontraditional" design choices might produce outlier results and graded those research designs accordingly.

One possible interpretation of this finding is that "run of the mill" specifications, which tended to confirm the Brady-Finnigan null results that all teams had reproduced, received the highest referee score, while the different specifications used by the ideological teams (in either direction) received worse scores. This does not necessarily imply that the run-of-the-mill approaches provide an unbiased estimate of the true parameter, as the unconventional approach may better approximate the data-generating process.

### Result limitations

We view the regressions reported in the text and in tables S2 to S10 in the Supplementary Materials as covering a large part of the universe of models that could logically represent the true (and unknown) data-generating process. Put differently, these regression models embody many of the analytical choices that researchers could justifiably and reasonably make in any examination of the experimental data. In a time of increasing concerns over the robustness of empirical social science research, the consistency of the substantive finding that ideology influences estimates of the impact of immigration suggests that our results qualify as evidence in secondary data analysis. Nevertheless, our examination of the experimental data has several limitations that are worth noting.

First, the original experiment was not designed to test (neither hypothetically nor based on a power analysis) the specific hypothesis examined in our paper: That ideology influences the production of research findings. The key motivation for the original experiment was to document the "hidden universe of data analysis," rather than identifying which factors create that universe. Although our statistical analysis reveals robust and strong differences in research findings across different kinds of researchers and/or teams, this limitation should not be ignored. Our findings may be considered a prior for future research as they provide exploratory evidence of the existence of ideological bias, but not a confirmatory test of any specific ideological bias hypothesis.

Second, although practically all our estimated ideology effects are statistically significant, the confidence intervals are often relatively large. As empirical social science shifts away from strict reliance on *P* values, a humbler interpretation of the evidence is not that ideology has a significant (rather than an insignificant) effect. Instead, it is that, under the assumption that the data-generating model implicitly adopted in our analysis is correctly specified, the ideology impact ranges from numerically small to large (though always in the same direction). This dispersion makes it difficult to draw a conclusion about the magnitude of this potential form of bias.

Third, few researchers revealed strong anti-immigration sentiments in the initial questionnaire (before any empirical analysis being done). This implies that relatively few teams can be classified as anti-immigration teams (only 9 teams of the 71 in the specifications that specifically allocate teams based on the immigration sentiment of their members). Our empirical analysis is then underpowered to identify a true effect between anti-immigration sentiments and the production of a negative impact of immigration on social cohesion. Nevertheless, our substantive conclusion is the same when the analysis instead compares the research findings produced by the 31 pro-immigration teams with the findings of all other teams (so that the comparison is based on larger sample sizes). The results of our study, therefore, may be more reliable when inferring the impact of pro-immigration sentiments on the production of research findings. Regardless, the key insight of the evidence is the same: Ideology plays a role in the production of research findings.

Fourth, it is possible that individual responses to the survey question on immigration sentiments were affected by social desirability bias and by concerns about being exposed as someone with strong anti-immigration sentiments (particularly in the academic environment where most of the individual researchers reside). Although the survey data collected from the individual researchers is anonymous publicly and within the group of the 158 researchers participating in the study, it was clear from the outset that the responses would be known to the principal investigators of the crowdsourced experiment. A sensible response to these social pressures implies that our sample of moderate or pro-immigration teams may contain some persons with anti-immigration sentiments. This contamination might help explain why the confidence intervals are so large if the hypothesis is true. Future studies analyzing similar questions could account for these social pressures during the design phase of the experiment, minimizing the potential bias introduced by misleading responses.

Last, although we have documented that ideology affects the choices researchers make in how they examine the same data, there is a fascinating and important question that remains unexplored: How exactly do researchers work their way through the maze of research design choices and end up with very specific specifications that lead to very specific empirical findings? The experimental data do not contain information on the extent to which researchers experimented with alternative specifications and why/when some specifications were discarded along the way. The careful observation and documentation of scientific workflows—i.e., the documentation of a researcher's "garden of forking paths"—may perhaps be the most exciting and promising avenue for future research in this area. In fairness, the mental gymnastics within a researcher's mind are inherently difficult to observe and document and might require an ideal experimental setting where every single step of an empirical exercise conducted by the researcher is tracked and recorded.

## DISCUSSION

We analyzed data generated by an experiment, where 71 research teams (comprising 158 individual researchers) used the same publicly available surveys to answer the same question: Does immigration

Downloaded from https://www.science.org on January 01, 2026

reduce the level of public support for the social programs that make up the welfare state? To anyone familiar with the mechanics of empirical work in the social and behavioral sciences, it is expected that there were as many estimates of this impact as there were alternative regression models (1253 to be exact). Each team had a way of selecting the sample for analysis, defining the key dependent and independent variables, and specifying the statistical analysis.

The data suggest that the team's pre-existing preference toward immigration restrictions plays a role in producing some of the observed differences. The estimated effect of immigration on social cohesion is more positive if the researchers are pro-immigration and more negative if the researchers are anti-immigration. Further, different combinations of five key design decisions between pro- and anti-immigration teams account for much of these results. These modeling choices also get lower "grades" from their peers than the specifications adopted by researchers who have moderate immigration sentiments. In other words, immigration ideology influences research quality in a way that happens to produce evidence that seemingly reflects the team's pre-existing ideology and is judged as worse suited to testing the hypothesis. Although our study is not the first to document a link between ideology and research findings, our analysis reveals that there is ideological bias across disciplines and even in a setting where the stakes are relatively low: The research was not part of a policy-oriented analysis and was not designed to culminate in stand-alone publications by the research teams.

One benign interpretation of our findings is that the time and effort devoted to research activities are limited resources, and researchers allocate their time and effort in the same way as everyone else. Namely, it is costly to develop an idea into a well-crafted empirical study, and researchers face many trade-offs when making this labor supply decision. Given the opportunity cost, it would be expected that once a researcher starts examining the data and finds a particularly appealing result that can be easily assembled into a compelling narrative, the researcher stops the empirical search for alternative stories. Metascience research reveals that this behavior is not uncommon (*13*, *33*). In this interpretation, researchers who are predisposed to favor a particular narrative about the impact of immigration begin to craft publishable results once the data seem to confirm an internally appealing story.

Even if this interpretation of the research process was correct, it highlights a problem. To precisely identify the causal impact of an exogenous shock on economic or social outcomes, the literature is increasingly dominated by policy evaluation studies. The shocks that are easy to find and examine empirically are typically shocks created by policy changes. This kind of policy-oriented research, however, may well attract a select group of researchers who really care about the consequences of the specific policies they plan to examine. Combined with the trade-offs in time and effort, the strong policy focus that motivates much of current empirical research could easily amplify the role of ideological bias in estimates of relevant parameters.

We recognize that the conclusions drawn from any analysis of this specific experiment are based on a moderately sized sample of 71 teams. Moreover, ideology was not a treatment variable randomly assigned in a controlled setting but simply a question asked to learn about researchers' background characteristics. This limits our ability to draw causal inferences. Nevertheless, comparing the research outcomes produced by teams who differ in their immigration sentiments across all our regression models (and hundreds of permutations of those models), we reject the statistical null hypothesis that ideology has a zero effect on the production of research findings in about 90% of the cases. We also recognize that although our results may not be generalizable, there probably are ideological differences among researchers examining any important issue in social policy, and our analysis may provide a better understanding of the evidence in those contexts. Our findings build a case for further development and testing of a theory of ideological bias in the production of research findings.

The ongoing revolution in artificial intelligence (AI) will likely influence the research process itself. The revolution introduces an additional source of ideological bias as the output of any AI algorithm may reflect the bias of its creators and certainly of the data that it is trained on (*34*, *35*). Moreover, the revolution may markedly lower the costs incurred by ideologically motivated researchers as they search across research designs for the model that confirms their priors. At the same time, however, just as an AI has already been trained to spot statistical errors (*36*), a future AI could theoretically be trained to identify ideological bias.

## METHODS

The Supplementary Materials give a detailed description of all variables used in the study and present additional tables and figures. Our workflow is online (https://github.com/nbreznau/ideology_specification).

The 2018 call for researchers to participate in the BRW experiment stated [SA, page 41 of (*20*)]: "We seek researchers to participate in a crowdsourced replication project on a high-profile social science question: **How does immigration shape public opinion? …** Participating researchers will (a) **replicate** and (b) **expand** a previously published **cross-national quantitative study**. We plan to distribute the results … and prepare them for publication in a high visibility social science journal. **All participants who complete the analytical tasks will be co-authors on the final paper …** We invite teams of 1-3 researchers to **independently analyze the data**" (bold in original).

The teams were instructed to first computationally reproduce results from Brady and Finnigan (*27*), a classic study of the hypothesis that correlated public attitudes toward the government provision of social programs with the level of immigration in 17 European countries using data from the 1996 and 2006 waves of the ISSP. The ISSP records attitudes toward government provision of social programs by asking: "Do you think it should or should not be the government's responsibility to…provide" these programs as old-age support, health, housing, and unemployment benefits [SA, page 7 of (*20*)]. The Brady-Finnigan study concluded that immigration did not affect public support for social programs.

After the reproduction of this null finding, the teams were then instructed to extend the analysis in the best way they saw fit. Each research team was given five waves of ISSP data (spanning 1985–2016), updated measures of immigration levels including percent foreign-born in the population and net immigration flows, and an updated array of country-level macro-indicator data. By the end of the project, BRW received results from 158 researchers in 71 teams.

Researcher information was collected by BRW in four survey waves, including educational background, familiarity with statistical methods, and prior experience in immigration or social policy research. A key question in these self-completed questionnaires (asked in the first wave, prior to any data analysis) ascertained the researcher's

Downloaded from https://www.science.org on January 01, 2026

stance toward immigration: "Do you think that, in your current country of residence, laws on immigration of foreigners should be relaxed or made tougher?" [SM, page 87 of (*20*)]. The researchers were given a seven-point scale (from 0 to 6) to answer the question, with the extremes being "immigration laws should be made tougher" and "immigration laws should be relaxed." Figure 1A illustrates the frequency distribution of the researchers' responses, which skew heavily toward a pro-immigration stance: Nearly half strongly support the proposal that immigration laws should be relaxed, choosing a "5" or a "6" as responses to the question, no researcher responded with a "0," and only 7.7% percent chose a "1" or a "2."

In their efforts to extend the Brady-Finnigan analysis, the teams estimated a total of 1253 regression models. The median team submitted 12 models, (10th/90th percentiles are 3 and 36, respectively). BRW converted the reported results into an "AME", giving the change in the probability of support for the government providing social programs resulting from a one percentage point increase in the immigrant population share. Figure 1B shows that although the AME frequency distribution clusters around zero, it also includes many numerically sizable and significant results. For example, the 10th percentile estimate is −0.071 (standard error = 0.019), and the 90th percentile is 0.052 (0.011). The 10th and 90th percentile estimates imply that social cohesion might be very responsive to immigration: A one percentage point change in the immigrant share (say, from 10 to 11% of the population) might reduce or increase the fraction of the population that supports social programs by −7 or +5 percentage points. Of course, the interesting question is whether the variation in the estimated effect of immigration is strictly random or can partly be attributed to pre-existing and observable researcher characteristics.

One measure of a team's immigration sentiment is the mean of the index illustrated in Fig. 1A across the (up to three) team members. It is obvious, however, that a simple averaging of this index may not completely capture the ideology of a particular team: Two three-person teams could both have a mean of 4.0, but this outcome could describe a team where all researchers responded with a "4" or a team where one researcher responded with a "6" and the other two responded with a "3." To allow for the possibility that some team members feel strongly about immigration (one way or the other) and to establish the robustness of our results, we use two additional measures of a team's ideology. The first is a two-variable vector giving the fraction of the team that is pro-immigration ("5" or "6") and the fraction that is anti-immigration ("1" or "2"). The second classifies teams into distinct categories (and this classification also helps visualize the evidence). A pro-immigration team is one where more than half the members respond with at least a "5" in the scale, and (given the relative rarity of anti-immigration sentiments) an anti-immigration team is one where at least one member responded with a "1" or "2". These definitions classify 31 of the 71 teams (43.7%) as pro-immigration and 9 teams (12.7%) as anti-immigration, with the remaining 31 (43.7%) as "moderate." Although this definition implies that relatively few teams are anti-immigration, we show below that the evidence is robust to alternative ways of describing immigration sentiments and that the evidence is robust to whether we measure these sentiments at the individual or team levels.

## Supplementary Materials

**This PDF file includes:**

Supplementary Text
Figs. S1 to S3
Tables S1 to S13

## REFERENCES


1. A. Manning, The elusive employment effect of the minimum wage. *J. Econ. Perspect.* **35**, 3–26 (2021).
2. D. Neumark, The econometrics and economics of the employment effects of minimum wages: Getting from known unknowns to known knowns. *Ger. Econ. Rev.* **20**, 293–329 (2019).
3. F. D. Blau, C. Mackie, Eds., *The Economic and Fiscal Consequences of Immigration* (National Academies Press, 2017).
4. C. Dustmann, U. Schönberg, J. Stuhler, The impact of immigration: Why do studies reach such different results? *J. Econ. Perspect.* **30**, 31–56 (2016).
5. J. Monras, Local adjustment to immigrant-driven labor supply shocks. *J. Hum. Cap.* **15**, 204–235 (2021).
6. M. Korbmacher, F. Azevedo, C. R. Pennington, H. Hartmann, M. Pownall, K. Schmidt, M. Elsherif, N. Breznau, O. Robertson, T. Kalandadze, S. Yu, B. J. Baker, A. O'Mahony, J. Ø.-S. Olsnes, J. J. Shaw, B. Gjoneska, Y. Yamada, J. P. Röer, J. Murphy, S. Alzahawi, S. Grinschgl, C. M. Oliveira, T. Wingen, S. K. Yeung, M. Liu, L. M. König, N. Albayrak-Aydemir, O. Lecuona, L. Micheli, T. Evans, The replication crisis has led to positive structural, procedural, and community changes. *Commun. Psychol.* **1**, 3 (2023).
7. M. L. Head, L. Holman, R. Lanfear, A. T. Kahn, M. D. Jennions, The extent and consequences of P-hacking in Science. *PLoS Biol.* **13**, e1002106 (2015).
8. A. Brodeur, N. Cook, A. Heyes, Methods Matter: p-hacking and publication bias in causal analysis in Economics. *Am. Econ. Rev.* **110**, 3634–3660 (2020).
9. A. Schneck, Are most published research findings false? Trends in statistical power, publication selection bias, and the false discovery rate in psychology (1975–2017). *PLOS ONE* **18**, e0292717 (2023).
10. A. S. Gerber, N. Malhotra, Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results? *Sociol. Method. Res.* **37**, 3–30 (2008).
11. C. Blanco-Perez, A. Brodeur, Publication bias and editorial statement on negative findings. *Econ. J.* **130**, 1226–1247 (2020).
12. J. Mervis, Why null results rarely see the light of day. *Science* **345**, 992–992 (2014).
13. L. K. John, G. Loewenstein, D. Prelec, Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* **23**, 524–532 (2012).
14. D. Kahneman, O. Sibony, C. R. Sunstein, *Noise. A Flaw in Human Judgement* (Little, Brown Spark, 2021).
15. C. J. Clark, B. M. Winegard, Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychol. Inq.* **31**, 1–22 (2020).
16. N. Honeycutt, L. Jussim, A model of political bias in social science research. *Psychol. Inq.* **31**, 73–85 (2020).
17. Z. Jelveh, B. Kogut, S. Naidu, Political language in Economics. *Econ. J.* **134**, 2439–2469 (2024).
18. S. I. Abramowitz, B. Gomes, C. V. Abramowitz, Publish or politic: Referee bias in manuscript review. *J. Appl. Soc. Psychol.* **5**, 187–200 (1975).
19. H. Finseraas, A. H. Midtbøen, K. Thorbjørnsrud, Ideological biases in research evaluations? The case of research on majority–minority relations. *Scand. Polit. Stud.* **45**, 370–381 (2022).
20. N. Breznau, E. M. Rinke, A. Wuttke, H. H. V. Nguyen, M. Adem, J. Adriaans, A. Alvarez-Benjumea, H. K. Andersen, D. Auer, F. Azevedo, O. Bahnsen, D. Balzer, G. Bauer, P. C. Bauer, M. Baumann, S. Baute, V. Benoit, J. Bernauer, C. Berning, A. Berthold, F. S. Bethke, T. Biegert, K. Blinzler, J. N. Blumenberg, L. Bobzien, A. Bohman, T. Bol, A. Bostic, Z. Brzozowska, K. Burgdorf, K. Burger, K. B. Busch, J. Carlos-Castillo, N. Chan, P. Christmann, R. Connelly, C. S. Czymara, E. Damian, A. Ecker, A. Edelmann, M. A. Eger, S. Ellerbrock, A. Forke, A. Forster, C. Gaasendam, K. Gavras, V. Gayle, T. Gessler, T. Gnambs, A. Godefroidt, M. Grömping, M. Groß, S. Gruber, T. Gummer, A. Hadjar, J. P. Heisig, S. Hellmeier, S. Heyne, M. Hirsch, M. Hjerm, O. Hochman, A. Hövermann, S. Hunger, C. Hunkler, N. Huth, Z. S. Ignácz, L. Jacobs, J. Jacobsen, B. Jaeger, S. Jungkunz, N. Jungmann, M. Kauff, M. Kleinert, J. Klinger, J.-P. Kolb, M. Kołczyńska, J. Kuk, K. Kunißen, D. K. Sinatra, A. Langenkamp, P. M. Lersch, L.-M. Löbel, P. Lutscher, M. Mader, J. E. Madia, N. Malancu, L. Maldonado, H. Marahrens, N. Martin, P. Martinez, J. Mayerl, O. J. Mayorga, P. M. Manus, K. M. Wagner, C. Meeusen, D. Meierrieks, J. Mellon, F. Merhout, S. Merk, D. Meyer, L. Micheli, J. Mijs, C. Moya, M. Neunhoeffer, D. Nüst, O. Nygård, F. Ochsenfeld, G. Otte, A. O. Pechenkina, C. Prosser, L. Raes, K. Ralston, M. R. Ramos, A. Roets, J. Rogers, G. Ropers, R. Samuel, G. Sand, A. Schachter, M. Schaeffer, D. Schieferdecker, E. Schlueter, R. Schmidt, K. M. Schmidt, A. Schmidt-Catran, C. Schmiedeberg, J. Schneider, M. Schoonvelde, J. Schulte-Cloos, S. Schumann, R. Schunck, J. Schupp, J. Seuring, H. Silber, W. Sleegers, N. Sonntag, A. Staudt, N. Steiber, N. Steiner, S. Sternberg, D. Stiers, D. Stojmenovska, N. Storz, E. Striessnig, A.-K. Stroppe, J. Teltemann, A. Tibajev, B. Tung, G. Vagni, J. Van Assche, M. van der Linden, J. van der Noll, A. Van Hootegem, S. Vogtenhuber, B. Voicu, F. Wagemans, N. Wehl, H. Werner, B. M. Wiernik, F. Winter, C. Wolf, Y. Yamada, N. Zhang, C. Ziller, S. Zins, T. Żółtak, Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203150119 (2022).


Downloaded from https://www.science.org on January 01, 2026

21. A. Alesina, E. Murard, H. Rapoport, Immigration and preferences for redistribution in Europe. *J. Econ. Geogr.* **21**, 925–954 (2021).
22. M. A. Eger, Even in Sweden: The effect of immigration on support for welfare state spending. *Eur. Sociol. Rev.* **26**, 203–217 (2010).
23. M. A. Eger, N. Breznau, Immigration and the welfare state: A cross-regional analysis of european welfare attitudes. *Int. J. Comp. Sociol.* **58**, 440–463 (2017).
24. A. W. Schmidt-Catran, D. C. Spies, Immigration and welfare support in germany. *Am. Sociol. Rev.* **81**, 242–261 (2016).
25. A. Alesina, E. Glaeser, *Fighting poverty in the US and Europe: A world of difference* (Oxford Univ. Press, 2004).
26. K. Auspurg, J. Brüderl, T. Wöhler, Does immigration reduce the support for welfare spending? A cautionary tale on spatial panel data analysis. *Am. Sociol. Rev.* **84**, 754–763 (2020).
27. D. Brady, R. Finnigan, Does immigration undermine public support for social policy? *Am. Sociol. Rev.* **79**, 17–42 (2014).
28. B. Burgoon, Immigration, integration, and support for redistribution in Europe. *World Polit.* **66**, 365–405 (2014).
29. H. Finseraas, Immigration and preferences for redistribution: An empirical analysis of European Survey Data. *Comp. Eur. Polit.* **6**, 407–431 (2008).
30. J. C. Garand, P. Xu, B. C. Davis, Immigration attitudes and support for the welfare state in the American mass public. *Am. J. Polit. Sci.* **61**, 146–162 (2017).
31. G. Dragolov, Z. S. Ignácz, J. Lorenz, J. Delhey, K. Boehnke, K. Unzicker, *Social Cohesion in the Western World: What Holds Societies Together: Insights from the Social Cohesion Radar* (Springer, 2016); https://link.springer.com/book/10.1007/978-3-319-32464-7.
32. A. Saltelli, A. Lachi, A. Puy, N. Breznau, "Global sensitivity analysis unveils the hidden universe of uncertainty in multiverse studies," in *Statistics for Innovation IV*, E. di Bella, V. Gioia, C. Lagazio, S. Zaccarin, Eds. (Springer Nature, 2025), pp. 299–304.
33. G. Gopalakrishna, G. T. Riet, G. Vink, I. Stoop, J. M. Wicherts, L. M. Bouter, Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLOS ONE* **17**, e0263023 (2022).
34. N. Breznau, Integrating computer prediction methods in Social Science: A comment on Hofman et al. (2021). *Soc. Sci. Comput. Rev.* **40**, 844–853 (2022).
35. M. Buyl, A. Rogiers, S. Noels, G. Bied, I. Dominguez-Catena, E. Heiter, I. Johary, A.-C. Mara, R. Romero, J. Lijffijt, T. De Bie, Large Language Models Reflect the Ideology of their Creators. arXiV: 2410.18417 [cs.CL] (2024) [accessed 9 November 2024]; http://arxiv.org/abs/2410.18417.
36. M. B. Nuijten, J. Wicherts, The effectiveness of implementing statcheck in the peer review process to avoid statistical reporting errors. PsyArXiv bxau9 [Preprint] (2023) [accessed 17 August 2023]; https://psyarxiv.com/bxau9/.



**Acknowledgments:** We are grateful to M. Amior, C. Amuedo-Dorantes, O. Ashenfelter, D. Brady, H. Cassidy, A. Edo, T. Emery, D. Hamermesh, T. Joyce, J. Llull, J. Monras, D. Rosenkranz, J. Stuhler, S. Trejo, R. VerBruggen, D. Weakleim, and C. Wlezien for valuable suggestions and discussions. **Funding:** This study was partially funded by a German Science Foundation grant, project number 464546557. **Author contributions:** Conceptualization: G.J.B. Data curation: N.B. and G.J.B. Formal analysis: G.J.B. and N.B. Investigation: G.J.B. and N.B. Methodology: G.J.B. and N.B. Project administration: G.J.B. Validation: G.J.B. and N.B. Visualization: N.B. and G.J.B. Writing—original draft: G.J.B. Writing—review and editing: N.B. and G.J.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. The data and code used in the analysis are available online at https://doi.org/10.5281/zenodo.17083885 (a packaged version of https://github.com/nbreznau/ideology_specification).

Submitted 11 June 2025
Accepted 19 November 2025
Published 1 January 2026
10.1126/sciadv.adz7173


Downloaded from https://www.science.org on January 01, 2026