Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-26-15

Geopolitics in the Evaluation of International Scientific Collaboration

Alexander Furnas
Northwestern University and IPR

Ruixue Jia
University of California, San Diego

Margaret Roberts
University of California, San Diego

Dashun Wang
Northwestern University

Version: June 11, 2026

DRAFT

Please do not quote or distribute without permission.

Abstract

This study provides evidence that geopolitical considerations systematically shape funding evaluations of international collaboration proposals. The authors examine this dynamic in the consequential context of U.S.–China collaboration. Across two large-scale randomized experiments with U.S. policymakers and U.S.-based scientists, the authors find substantial and consistent penalties for proposals involving China-based collaborators. Policymakers express much greater unconditional support for proposals with Germany-based collaborators than for otherwise identical proposals with China-based collaborators (68% vs. 28%). Crucially, this penalty is not confined to policymakers: scientists themselves exhibit a sizeable 18 percentage-point gap (48% vs. 30%), despite professional expectations of merit-based evaluation. Much of the difference reflects a shift from unconditional to conditional approval rather than outright rejection. These penalties are remarkably consistent across scientific fields and respondent characteristics, with little evidence of heterogeneity, indicating that they reflect geopolitical rather than domain-specific concerns. Overall, the findings suggest that geopolitics influences gatekeeping judgments in government funding, with broad implications for peer review, scientific norms, and the future of international collaboration in an era of intensifying geopolitical competition.

Introduction

Scientific research has long been guided by ideals of openness, universalism, international collaboration, and the free exchange of knowledge across borders [1–3]. Many of the world's most significant scientific advances have emerged from partnerships that transcend national boundaries—like the discovery of the Higgs Boson at CERN or the Human Genome Project—contributing to a vision of science as a universal enterprise dedicated to the collective pursuit of knowledge [1, 4, 5]. Yet science is also a public good, supported by national funding and intertwined with national priorities—whether advancing economic growth, technological leadership, or national security [6–8]. This duality creates a fundamental tension: as nations navigate geopolitical competition, the very openness that has fueled scientific progress comes under pressure, raising questions about how global collaboration aligns with national interests.

Nowhere is this tension more visible than in the evolving relationship between the United States and China. Over the past two decades, U.S.-China scientific collaborations have produced some of the most impactful research worldwide [9], spanning fields from astronomy to cancer research [10]. At the same time, rising concerns over intellectual property, surveillance, human rights, and national security have led to increased scrutiny and restrictions on such partnerships [11–16], and driving political debate over the risks and benefits of scientific collaboration with China [17, 18]. This contested terrain reflects broader anxieties about the role of science in an era of geopolitical rivalry, where scientific partnerships are viewed not only as vehicles for discovery but also as conduits for strategic influence.

While prior research has documented how geopolitical events may disrupt scientific productivity and international collaboration [19–25], little is known about how geopolitical considerations influence the evaluation of science itself. In particular, it remains unclear whether—and to what extent—the decision-makers who govern scientific funding, including policymakers and scientists, incorporate geopolitical factors into their judgments regarding funding decisions. Addressing this question is critical for understanding how global scientific cooperation is shaped, sustained, or constrained under conditions of geopolitical competition.

Here we investigate this question by conducting two large-scale randomized experiments across two critical gatekeeping communities: U.S. policymakers and active U.S.-

based scientists. Conceptually, our focus is narrow but consequential. We examine how the national affiliation of a proposed research collaborator—specifically China versus a close U.S. ally—causally affects individual evaluative judgments in a government funding context, holding constant project content, quality signals, and review criteria. We do not attempt to measure geopolitical attitudes directly, nor do we estimate historical change. Rather, our design is cross-sectional – we identify how current geopolitical relations, as reflected in collaborator nationality, shape contemporaneous funding evaluations among key scientific gatekeepers, and does not require a historical baseline to interpret. Respondents were presented a description of a hypothetical research proposal submitted to the NSF (see SM). The research was explicitly described as unclassified and not subject to export control. We randomly assigned two key variables in our survey: (1) whether the research proposal involved collaboration between scientists in the U.S. and Germany or between scientists in the U.S. and China, and (2) the scientific field of the collaboration (batteries, biotechnology, robotics, or environmental science).[1]

Our results reveal clear and consistent patterns: Both policymakers and scientists exhibit significant reluctance to unconditionally support collaborations with scientists in China compared to otherwise identical collaborations with scientists in Germany, even when proposals are described as meeting peer-review criteria. Further, the treatment effects we observe are pronounced and exhibit remarkably little heterogeneity, as proposals featuring China-based collaborators are universally penalized to a large margin. Importantly, rather than rejecting such collaborations outright, many respondents expressed conditional support, especially when safeguards are in place to address concerns about national security, surveillance, and human rights.

Existing work documents downstream consequences of geopolitical tension—such as changes in collaboration networks, citation flows, and productivity—but cannot distinguish whether these patterns arise from researcher behavior, institutional constraints, or evaluative gatekeeping. Our contribution is to isolate, with experimental control, how collaborator nationality alone shapes funding judgments before any collaboration occurs, and to do so among both policymakers and scientists who routinely act as evaluators. This identifies a distinct mechanism—ex ante gatekeeping—that cannot be inferred from

[1] We chose these fields because they spann a range of critical technology areas where U.S. and China hold different comparative advantages, but with varying degrees of relevance to the military and commercial sectors

observational bibliometric data alone.

By showing that geopolitical considerations shape how science is evaluated, even by scientists themselves, our findings reveal how international politics penetrate the gatekeeping mechanisms of scientific knowledge. Scientists' attitudes on this issue, much like those of policymakers, are particularly striking and challenge the common assumption that scientific judgment operates independently of geopolitical concerns. These results carry important implications for the governance of international scientific collaboration, the resilience of global scientific networks, and the future of scientific openness in an era of intensifying geopolitical competition.

Results

Our first experiment includes 1,249 participants, comprising current and former U.S. government officials responsible for trade, development, or national security policy, think tank members specializing in international affairs, and international relations (IR) scholars. To ensure practical relevance, we partnered with the Teaching, Research, and International Policy (TRIP) lab, known for its extensive experience conducting foreign policy surveys [26, 27].[2]

In our design, we distinguish between unconditional funding, conditional funding, and rejection because these categories map onto distinct policy actions: full approval, approval with contingent on additional safeguards, and denial. While real-world awards often involve administrative conditions, the conditions offered here capture additional, country-contingent requirements beyond standard compliance, allowing us to isolate how collaborator nationality changes the stringency of evaluation.

Figure 1A illustrates a stark difference in policymakers' support for international collaborations. While 68.2% support funding for U.S.-Germany collaborations, only 27.6% offer similar support for otherwise identical collaborations involving China. This indicates policymakers are more than twice as likely to unconditionally support collaborations with Germany-based researchers compared to their China-based counterparts, despite both proposals having cleared peer review (categorical: $F(2.00, 2496) = 108.2$, $p < .001$; binarized unconditional support: $F(1, 1248) = 205.86$, $p < .001$; Rao–Scott adjusted Chi-squared

[2]This study was approved by Institutional Review Boards at William and Mary and UC San Diego.

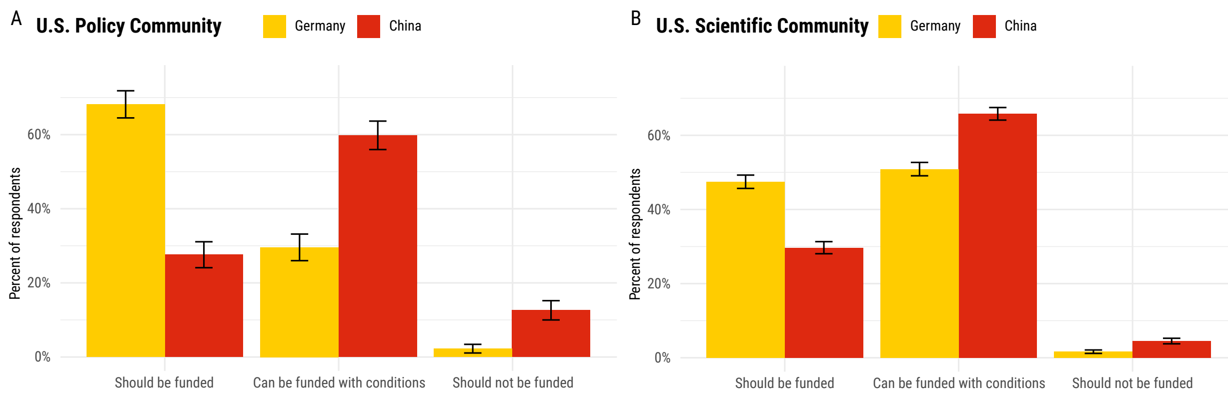


Figure 1: (A) Support for U.S.-China Collaborations vs. U.S.-Germany Collaborations in the U.S. Policy Community. (B) Support for U.S.-China Collaborations vs. U.S.-Germany Collaborations among U.S.-based Scientists. Error bars indicate 95% CIs.

tests). Importantly, the China penalty manifests primarily as heightened conditionality rather than categorical rejection, indicating caution and risk management rather than blanket opposition. This disparity holds across all four fields of research we examined, indicating that the gap reflects geopolitical rather than domain-specific concerns(see SM).

This pronounced disparity, though notable, may partly reflect policymakers' inherently political roles and sensitivity toward geopolitics. In contrast, science funding is traditionally guided by scientific experts who adhere to professional norms emphasizing objectivity and meritocracy [7]. Therefore, one might anticipate the differences we observed for the policy community may be greatly attenuated in the scientific community, as scientists evaluate otherwise identical collaborations based solely on national origin. To explore this, we conduct Study 2, repeating the same experimental setup with active U.S.-based scientists.

We identified corresponding authors from articles published in Web of Science-indexed journals [28] between January 2022 and September 2024. From this dataset, we randomly selected 245,000 U.S.-affiliated scholars. After cleaning inactive and duplicate email addresses, we distributed 218,195 invitations, receiving 9,830 completed responses, representing a response rate of approximately 4.5%, consistent with prior survey studies [e.g. 29]. Following screening for informed consent, U.S. location, and active research status, we finalized an analysis sample of 7,154 scientists.[3] We constructed post-stratification weights using participants' h-index, publication history, and research field to enhance rep-

[3] Effective sample size post-weighting: 5,912 [30].

resentativeness (detailed weighting methods and covariate balance statistics provided in SM). All results presented incorporate these weights; unweighted analyses yield substantively identical conclusions (see SM).

Overall, scientists exhibit a similar, though somewhat narrower, disparity in their evaluations (Fig. 1(B)). Specifically, 47.5% support unconditional funding for German collaborations compared to 29.7% for identical collaborations involving scientists in China. Thus, scientists are roughly 60% more inclined to endorse funding collaborations with scholars in Germany—a substantial difference given typical baseline funding rates at NSF (categorical: $F(2.00, 14{,}302.31) = 108.3$, $p < .001$; binarized unconditional support: $F(1, 7153) = 200.14$, $p < .001$; Rao–Scott adjusted Chi-squared tests). These results are largely consistent across the fields of research presented in the vignette, although they are slightly less pronounced in environmental science (see SM).

However, like policymakers, scientists predominantly prefer conditional support rather than outright rejection of collaborations with scientists in China (Fig. 2). Indeed, 65.8% of scientists favor conditional funding for China-based collaborations, primarily emphasizing conditions related to national security, surveillance, and human rights concerns. In contrast, conditional support for collaborations with scientists in Germany is lower, at 50.9%. Frequently imposed conditions include ensuring collaborators have no military affiliations (53.1%) and prohibiting surveillance applications (49.1%), significantly higher than conditions set for German collaborations (35%). Scientists also more often require collaborations with scientists in China to explicitly benefit global common goods such as climate or health research.

Germany serves here as a theoretically meaningful benchmark: a close U.S. ally with strong scientific capacity and minimal security concerns. Our estimates therefore identify reluctance toward China relative to a best-case collaboration partner, not relative to a hypothetical neutral country—an important boundary condition for interpretation.

Subgroup analyses reveal more nuanced variations. There is remarkably little heterogeneity in treatment effects in the policy community, although the disparity between evaluations of collaborations with China-based and Germany-based scientists is somewhat smaller among IR scholars than think tankers or policymakers (see SM). Turning to the scientist sample, senior and more established scientists (with higher h-index) exhibit slightly greater reluctance toward unconditional funding of collaborations with China compared to junior researchers. At the same time, these subgroup differences are minor in magni-

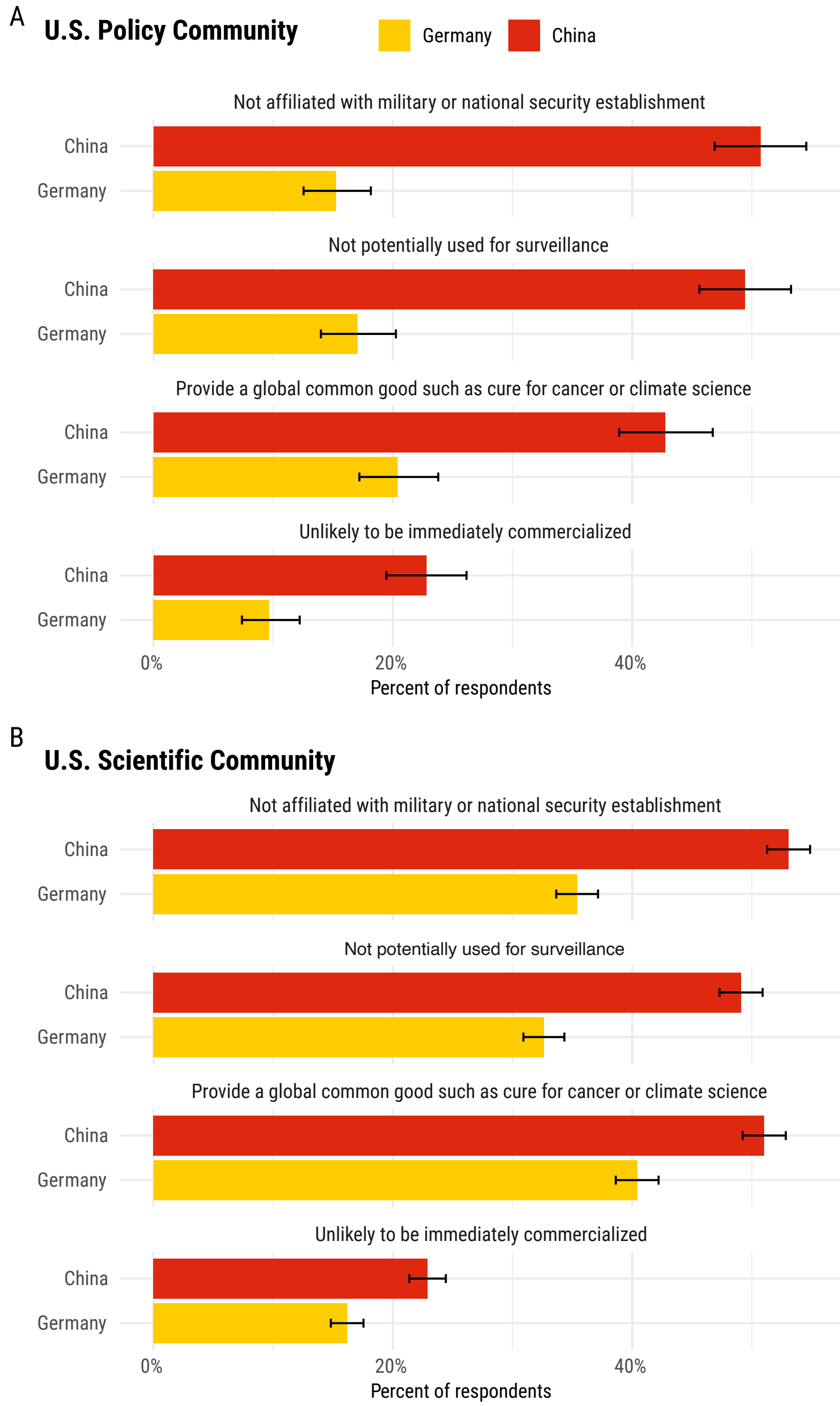


Figure 2: (A) Conditional support for U.S.-China Collaborations vs. U.S.-Germany Collaborations in the U.S. Policy Community. (B) Conditional support for U.S.-China Collaborations vs. U.S.-Germany Collaborations among U.S.-based Scientists. Error bars indicate 95% bootstrapped CIs.

tude compared to the overall, pervasive disparity in support between Germany and China collaborations, underscoring the robustness of geopolitical considerations influencing scientific decision-making (additional details provided in SM).

Taken together, results from both surveys closely mirror each other, consistently highlighting caution toward U.S.-China scientific collaborations among both policymakers and scientists. The effect sizes we observe are consistent and substantial in magnitude. This alignment suggests geopolitical tensions permeate beyond purely political spheres, potentially exerting a broader chilling effect on international collaborations involving China. Notably, these patterns persist regardless of the specific research fields explored in our experiments, highlighting that geopolitical factors, rather than domain-specific concerns, are largely responsible for the observed disparities.

Nevertheless, noteworthy differences emerge between the policy and scientific communities. Scientists are generally more open than policymakers to conditional funding, regardless of collaborator nationality. For China-based collaborations, nearly two-thirds of scientists (65.8%) support funding under conditions, compared to about three-fifths of policymakers (59.8%). The gap is even starker for Germany-based collaborations: roughly half of scientists (50.9%) endorse conditional funding, compared to less than one-third of policymakers (29.6%). Furthermore, categorical rejection is markedly higher among policymakers: 12.6% outright oppose funding collaborations with China, roughly six times the baseline rejection rate for Germany (2.3%). In comparison, scientists show a smaller disparity, with 4.5% rejecting collaborations with scientists in China compared to a baseline of 1.9% for collaborations with scientists in Germany. These results demonstrate that while geopolitical considerations significantly influence both communities, policymakers exhibit more pronounced reluctance toward collaborations involving China.

Alternative Interpretations

Taken together, these experiments provide evidence that geopolitical considerations shape the evaluation of scientific research. Could alternative explanations account for our main findings? One possibility is cultural familiarity: American scientists may be less familiar with, or less inclined to collaborate with, China-based researchers for reasons unrelated to geopolitics. However, the evidence does not support this interpretation.

First, the stability of the estimated effects across scientific fields and respondent characteristics constrains explanations that rely primarily on idiosyncratic heuristics, field-specific risk, or differential familiarity. If cultural distance or collaboration experience were the primary drivers, we would expect substantial heterogeneity across disciplines or researcher backgrounds. Instead, the near-universality of the effects makes such accounts difficult to reconcile with the data.

Second, additional analyses show that the patterns are remarkably similar among scientists who have previously collaborated with China-based researchers and those who have not. Moreover, the disparity persists even among Asian scientists (see SM). These findings directly undermine explanations based on cultural unfamiliarity or reluctance rooted in ethnic or social distance.

Another potential critique is that lower support for collaboration with Chinese scientists may have predated the recent escalation of geopolitical tensions. While we do not have direct evidence on individual attitudes prior to this period, existing research on scientific collaboration patterns consistently documents a sustained rise in U.S.–China collaborations through the mid-2010s, followed by a sharp decline beginning around 2018 [19–25]. In addition, the specific conditions highlighted in our survey—such as concerns about national security and surveillance—closely mirror the criteria that have become salient in scientific evaluations in the current geopolitical environment. Together, these patterns are more consistent with a geopolitically driven shift than with long-standing cultural frictions.

Discussions and Conclusions

By conducting two large-scale survey experiments across critical gatekeeping communities—U.S. policymakers and active U.S.-based scientists—we uncover striking patterns indicating that the country of a potential collaborator significantly influences funding decisions for international collaborations. Both policymakers and scientists were consistently less supportive of proposals involving China-based collaborators than of otherwise identical proposals involving German collaborators, despite being explicitly prompted that proposals have passed peer review and were unclassified and not subject to export control. The treatment effects we observe are substantial, and exhibit remarkably little heterogeneity, with proposals featuring China-based collaborators universally penalized.

At a conceptual level, our results highlight an important yet underappreciated tension within science policy. Science is often idealized as a universal, open, and politically neutral endeavor aimed at the advancement of collective knowledge. However, in practice, science remains a national enterprise supported by taxpayer funds and is often strategically leveraged to advance national interests. Our study suggests that geopolitical competition significantly shifts the balance toward nationalistic considerations, even among scientists themselves.

Perhaps most strikingly, this pattern extended significantly into the scientific community. Science funding decisions are traditionally driven by peer-review processes emphasizing objectivity and meritocratic evaluation. Recent work shows that political control substantially influences the amount of science funding [31], though processes like peer review can buffer against such swings. The substantial differences we observe among scientists based solely on the country of the collaborator, however, is striking and normatively salient. One might anticipate policymakers, inherently sensitive to national strategic interests, to exhibit pronounced differences; however, scientists' judgments were often presumed insulated by professional norms of objectivity and universalism. The fact that scientists themselves show significant geopolitical sensitivity raises important questions about the extent to which such considerations may be more pervasive and deeply embedded in the scientific evaluation process than previously recognized, with implications for not only science evaluation but also the norms of science.

While these findings may resonate with policymakers who have urged universities to treat U.S.–China collaboration as a strategic risk domain, they also raise a more basic concern: what criteria do scientists apply when geopolitical considerations enter evaluative judgments that are ostensibly merit-based? Most scientists are not trained to assess national security, human rights, or geopolitical risk, so their judgments may reflect informal heuristics rather than transparent, consistently applied standards, creating a tension between legitimate external constraints and the integrity of scientific evaluation.

Existing institutional guidance illustrates the kinds of considerations that can legitimately generate conditionality. For example, some U.S. universities advise PIs to scrutinize potential military ties and avoid projects that could contribute to surveillance or the suppression of human rights.[4] In our data, these concerns appear salient: respondents disproportionately impose security-related conditions on China-based collaborations.

At the same time, formal security guidance cannot fully explain for the patterns we

[4] https://global.mit.edu/wp-content/uploads/2022/11/FINALUniversity-Engagement-with-China_An-MIT-Approach-Nov2022.pdf

observe. Scientists not only apply security- and surveillance-related conditions more frequently to China than to Germany, but also more often require that China collaborations serve a global common good and are more likely to reject them outright, suggesting that proposals involving China are sometimes held to a higher bar or evaluated in a more country-contingent manner than proposals involving a close ally.

Overall, these findings invite critical reflection about whether and how such geopolitical factors extend into real-world evaluative contexts. Although our experiments rely on hypothetical proposals, real-world funding decisions ultimately aggregate individual judgments of precisely this kind. While deliberation, accountability, and institutional rules may attenuate or amplify individual judgments in this evaluative task, they cannot operate independently of them. Our results therefore identify a necessary input to real gatekeeping processes, even if they do not exhaustively characterize those processes, suggesting fruitful avenues for further work exploring how these attitudes may permeate actual peer-review processes, grant decisions, personnel evaluations, and even editorial judgments at scientific journals.

Further, our design focused on U.S. government funding, an especially consequential context given its role in setting national research priorities and its sensitivity to geopolitical concerns. Future work should examine whether similar patterns extend to other evaluative contexts, including those less directly tied to national policy priorities, especially considering well-documented bias against East Asian authors across multiple evaluative settings [32–34]. Such biases could disadvantage researchers affiliated with geopolitically sensitive nations, systematically affecting career trajectories, research funding, and scientific recognition. While survey experiments like ours provide robust causal evidence, extending these findings to real-world decision-making contexts remains an important but challenging goal, given the ethical and practical constraints on controlled field experimentation in high-stakes grant review.

In an era of intensified geopolitical competition, these findings show that collaborator nationality systematically shapes funding evaluations at the point of gatekeeping, highlighting a fundamental tension between scientific universalism and national strategic concerns. Addressing this tension, by protecting the integrity, objectivity, and internationalism of science while safeguarding legitimate national security and ethical considerations, will be critical for sustaining scientific progress and effective global cooperation on shared challenges. Funding agencies and scientific institutions might benefit from explicitly considering these issues, exploring measures that could help safeguard fairness and objectivity without compromising legitimate national security or ethical considerations.

Data Availability Statement Anonymized data required to reproduce the results will be made available upon publication.

References

[1] Robert K Merton. The sociology of science: Theoretical and empirical investigations. University of Chicago press, 1973.

[2] National Academies of Sciences, Medicine, Global Affairs, Board on Research Data, and Committee on Toward an Open Science Enterprise. Open science by design: Realizing a vision for 21st century research. 2018.

[3] Caroline S Wagner, Han Woo Park, and Loet Leydesdorff. The continuing growth of global cooperation networks in research: A conundrum for national governments. PloS one, 10(7):e0131816, 2015.

[4] Dashun Wang and Albert-László Barabási. The science of science. Cambridge University Press, 2021.

[5] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. Science of science. Science, 359(6379):eaao0185, 2018.

[6] Yian Yin, Yuxiao Dong, Kuansan Wang, Dashun Wang, and Benjamin F Jones. Public use and public funding of science. Nature human behaviour, 6(10):1344–1350, 2022.

[7] Vannevar Bush. Science, The Endless Frontier. U.S. Government Printing Office, Washington, D.C., 1945. Report to the President.

[8] Benjamin F Jones. Science and innovation: The under-fueled engine of prosperity. Rebuilding the Post-Pandemic Economy, ed. Melissa S. Kearney and Amy Ganz (Washington DC: Aspen Institute Press, 2021), 2021.

[9] Qingnan Xie and Richard B Freeman. Creating and connecting us and china science: Chinese diaspora and returnee researchers. Technical report, National Bureau of Economic Research, 2023.

[10] James Mitchell Crow. Us–china partnerships bring strength in numbers to big science projects. Nature, 603(7900):6–8, 2022.

[11] Eileen Guo, Aloe Jess, and Hao Karen. The us crackdown on chinese economic espionage is a mess. we have the data to show it. MIT Technology Review, 2021.

[12] Rob Portman and Tom Carper. Threats to the us research enterprise: China's talent recruitment plans. Online: https://www. hsgac. senate. gov/imo/media/doc/2019-11-18% 20PSI% 20Staff% 20Report, 2019.

[13] Jeffry Mervis. Pall of suspicion. Science, 379(6638):1180–1184, 2023.

[14] Jeffrey Stoff and Glenn Tiffert. Eyes wide open: Ethical risks in research collaboration with china. Technical report, Hoover Institution, 2021.

[15] Stuart Anderson. Biden keeps costly trump visa policy denying chinese grad students. Forbes. Retrieved from https://www. forbes. com/sites/stuartanderson/2021/08/10/biden-keeps-costly-trump-visa-policy-denying-chinese-gra 2020.

[16] David Shepardson. Us finalizes rules to prevent china from benefiting from $52 billion in chips funding. Reuters. Retrieved from https://www.reuters.com/technology/us-finalizes-rules-prevent-china-benefiting-52-bln-chips-fun 2023.

[17] JJ Lee and X Li. Racial profiling among scientists of chinese descent and consequences for the us scientific community. Committee of 100, 2021.

[18] Agathe Demarais. What does 'de-risking' actually mean? Foreign Policy, 2023.

[19] George J Borjas and Kirk B Doran. The collapse of the soviet union and the productivity of american mathematicians. The Quarterly Journal of Economics, 127(3): 1143–1203, 2012.

[20] Fabian Waldinger. Peer effects in science: Evidence from the dismissal of scientists in nazi germany. The review of economic studies, 79(2):838–861, 2012.

[21] Petra Moser, Alessandra Voena, and Fabian Waldinger. German jewish émigrés and us invention. American Economic Review, 104(10):3222–3255, 2014.

[22] Ruixue Jia, Margaret E Roberts, Ye Wang, and Eddie Yang. The impact of us–china tensions on us science: Evidence from the nih investigations. Proceedings of the National Academy of Sciences, 121(19):e2301436121, 2024.

[23] Philippe Aghion, Céline Antonin, Luc Paluskiewicz, David Strömberg, Xueping Sun, Raphaël Wargon, and Karolina Westin. Does chinese research hinge on us coauthors? evidence from the china initiative. 2023.

[24] Yu Xie, Xihong Lin, Ju Li, Qian He, and Junming Huang. Caught in the crossfire: Fears of chinese–american scientists. Proceedings of the National Academy of Sciences, 120 (27):e2216248120, 2023.

[25] Moxin Li and Yang Wang. Influence of political tensions on scientific productivity, citation impact, and knowledge combinations. Scientometrics, 129(4):2337–2370, 2024.

[26] Paul C Avey, Irene Entringer Garcia Balnes, Shauna N. Gillooly, Susan Peterson, Ryan Powers, Andrew Shaver, and Michael J. Tierney. Trip policymaker survey report. Teaching, Research and International Policy (TRIP) Project Global Research INstitute. Available at https://trip.wm.edu., 2023.

[27] Paul C Avey, Michael C Desch, Eric Parajon, Susan Peterson, Ryan Powers, and Michael J Tierney. Does social science inform foreign policy? evidence from a survey of us national security, trade, and development officials. International Studies Quarterly, 66(1), 2022.

[28] Philippe Mongeon and Adèle Paul-Hus. The journal coverage of web of science and scopus: a comparative analysis. Scientometrics, 106:213–228, 2016.

[29] Jian Gao, Yian Yin, Kyle R Myers, Karim R Lakhani, and Dashun Wang. Potentially long-lasting effects of the pandemic on scientists. Nature communications, 12(1): 6188, 2021.

[30] Leslie Kish. Sampling organizations and groups of unequal sizes. American sociological review, pages 564–572, 1965.

[31] Alexander C Furnas, Nic Fishman, Leah Rosenstiel, and Dashun Wang. Partisan disparities in the funding of science in the united states. Science, 389(6766):1195–1200, 2025.

[32] Hao Peng, Karim Lakhani, and Misha Teplitskiy. Acceptance in top journals shows large disparities across name-inferred ethnicities. 2021.

[33] Jackson G Lu, Richard E Nisbett, and Michael W Morris. Why east asians but not south asians are underrepresented in leadership positions in the united states. Proceedings of the national academy of sciences, 117(9):4590–4600, 2020.

[34] Jackson G Lu, Richard E Nisbett, and Michael W Morris. The surprising underperformance of east asians in us law and business schools: The liability of low assertiveness and the ameliorative potential of online classrooms. Proceedings of the National Academy of Sciences, 119(13):e2118244119, 2022.

Supplementary Materials: Geopolitics in the Evaluation of International Scientific Collaboration

1. INTRODUCTION

2. STUDY 1 - POLICY COMMUNITY

The survey for Study 1 was conducted in collaboration with the Teaching, Research, and International Policy (TRIP) lab, a well-established research lab at William and Mary with a two-decade history of conducting surveys on foreign policy issues [1, 2]. The survey sample includes 254 individuals who currently hold or have previously held positions within U.S. government agencies responsible for trade, development, or national security policy; 227 members of think tanks focused on international affairs; and 768 faculty members from political science departments or professional schools with a focus on international relations.[1]

2.1. Experimental Design

The experimental setup presented respondents with a hypothetical research proposal submitted to the NSF (Fig. S1). The research was explicitly described as unclassified and not subject to export control. We randomized two key variables: (1) whether the research proposal involved collaboration between scientists in the U.S. and Germany or between scientists in the U.S. and China, and (2) the scientific field of the collaboration. Respondents were asked under what conditions they believed the proposal should be funded. Respondents who selected the option to conditionally support funding were then prompted to specify any conditions they deemed necessary for the funding to be granted. They could choose from the following options: (1) the collaborator should not have any affiliation with the military or national security establishment, (2) the research should have no potential use in surveillance activities, (3) the research should contribute to a global common good, such as curing cancer or advancing climate science, and (4) the research should be unlikely to be immediately commercialized. These options were designed to assess the relative importance of security and commercial considerations in the respondents' decision-making processes.

The introductory language references concerns about foreign influence to situate respondents in a realistic policy environment. Because this framing is held constant across treatment conditions, it cannot account for the observed China–Germany differences, though it may raise baseline caution overall. We therefore interpret our estimates as relative effects within a security-salient context, consistent with real-world funding deliberations.

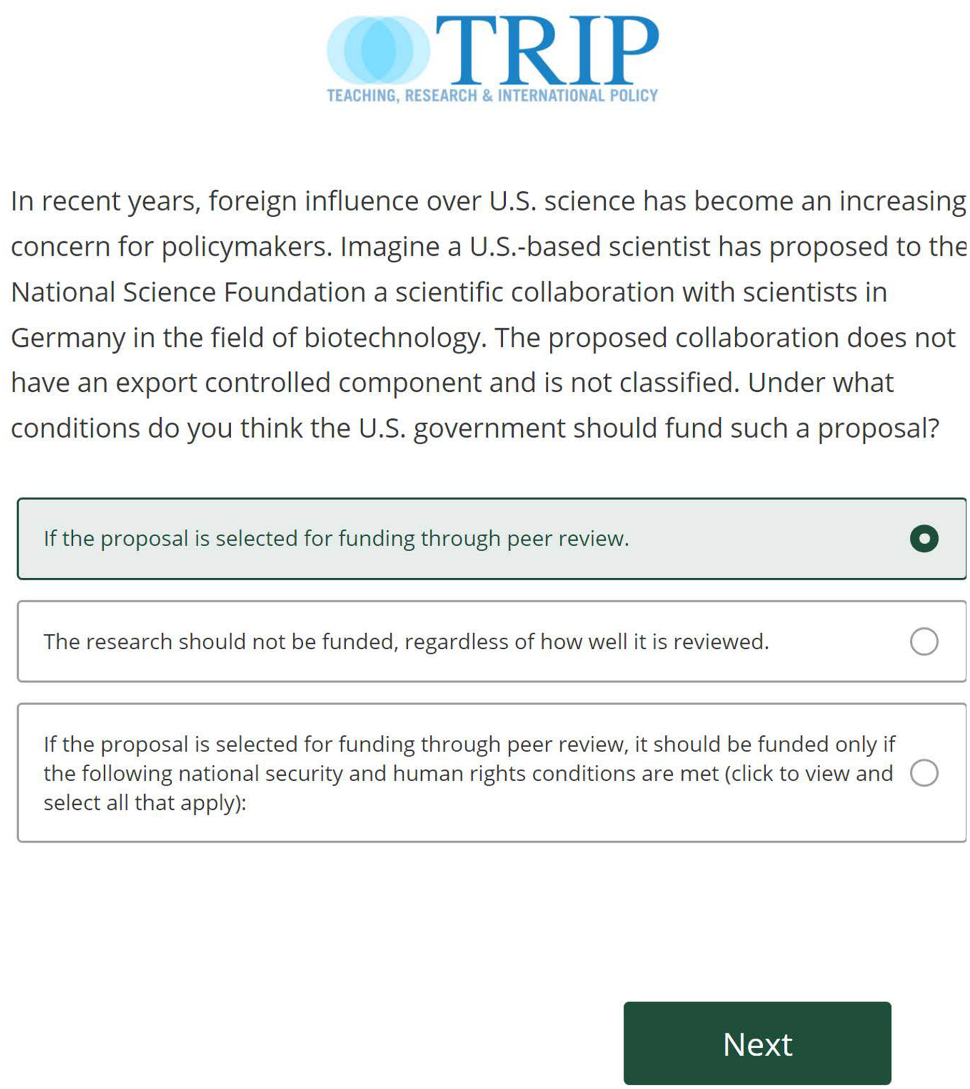
2.2. Validating Our Randomization Design

Figure S2 provides an overview of the demographic characteristics of the survey participants, including gender, race, age, political party affiliation, and ideological leaning on economic and social issues. The demographic composition of our sample closely mirrors that of the broader policy community. Crucially, the attributes of the participants are evenly distributed across the different experimental conditions, which confirms the effectiveness of our randomization process.

2.3. Supplementary Results

These results are largely consistent across the three groups of participants, with support considerably lower for U.S.-China science collaborations than U.S.-Germany collaborations in all cases. However, IR scholars show greater support for U.S.-China collaborations compared to policymakers and think tank members ($F(4.00, 4992) = 4.318$, $p < .01$; Rao–Scott adjusted Chi-squared test in the three-category outcome and $F(2, 2496) = 7.485$, $p < .001$ in binarized, unconditional funding support response variable; Rao–Scott adjusted Chi-squared tests). The likelihood of an IR

[1] This study was approved by the Institutional Review Boards at William and Mary and UC San Diego.


TRIP
TEACHING, RESEARCH & INTERNATIONAL POLICY

In recent years, foreign influence over U.S. science has become an increasing concern for policymakers. Imagine a U.S.-based scientist has proposed to the National Science Foundation a scientific collaboration with scientists in Germany in the field of biotechnology. The proposed collaboration does not have an export controlled component and is not classified. Under what conditions do you think the U.S. government should fund such a proposal?

- If the proposal is selected for funding through peer review.
- The research should not be funded, regardless of how well it is reviewed.
- If the proposal is selected for funding through peer review, it should be funded only if the following national security and human rights conditions are met (click to view and select all that apply):

Next

Fig. S1. Survey for Study 1, First Page

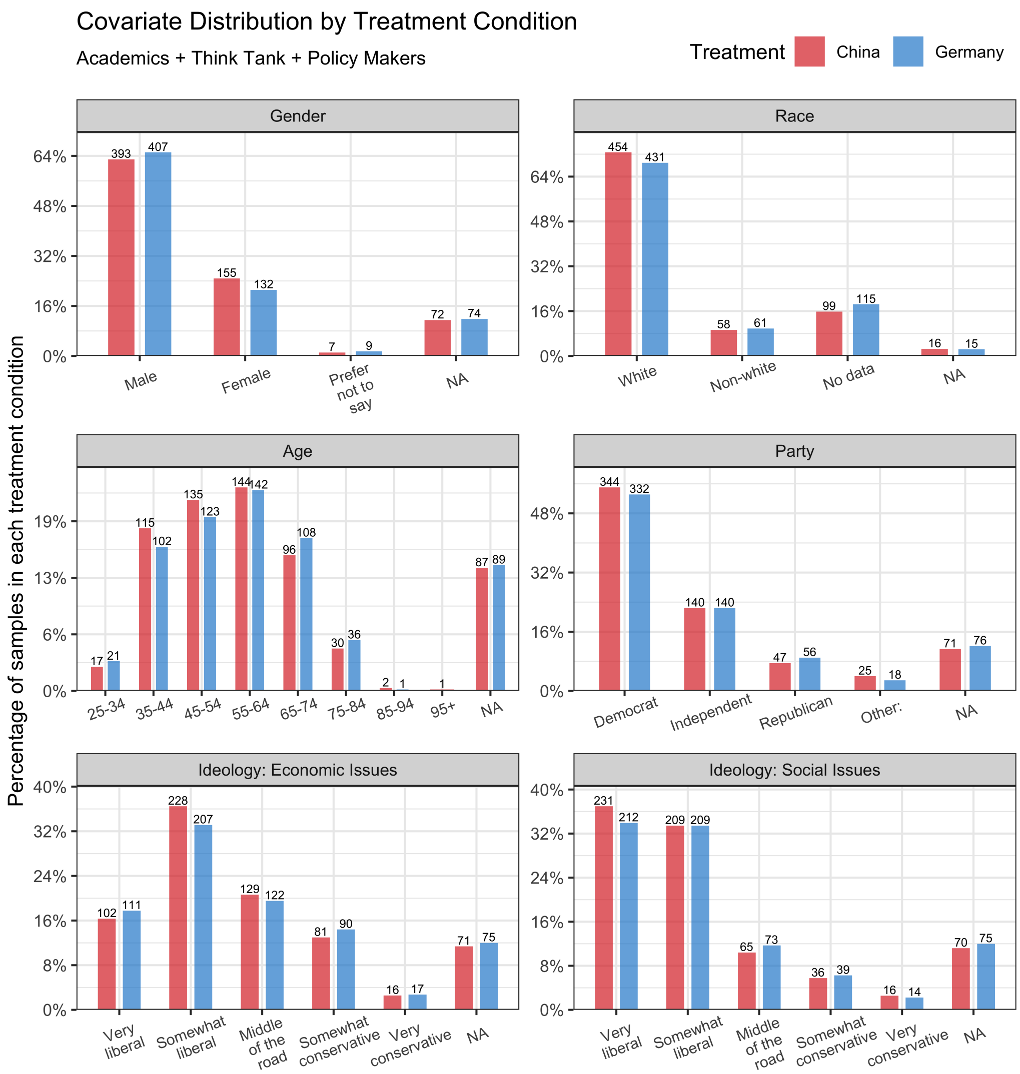


Fig. S2. Personal Characteristics across Questions

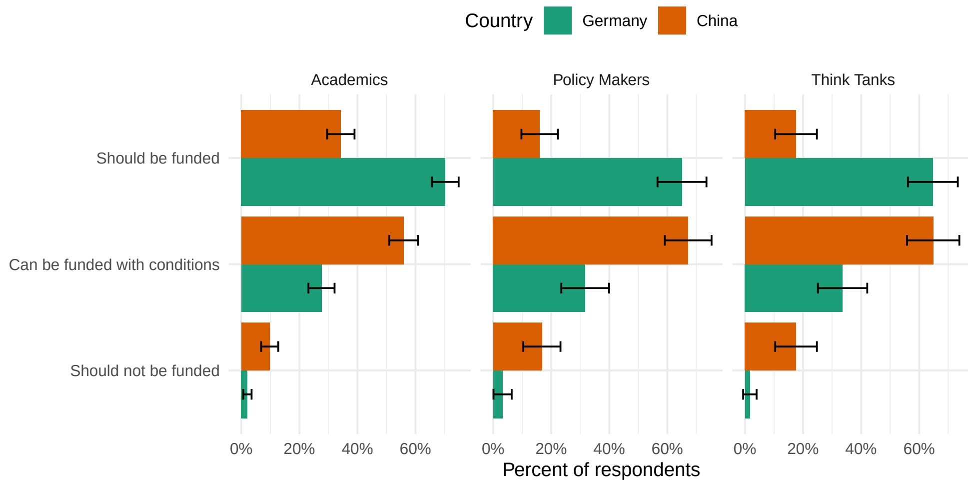


Fig. S3. Support by type of policy community respondent.

scholar supporting funding of a China collaboration based only on peer review is 34.28%, which is double the rate for policymakers at 16.03% and think tank members at 17.59%

In Figure S3 we present the percent of respondents selecting each response option broken out separately by type of respondent in the TRIP survey.

We find large differences in level of support for collaboration with China and collaboration with Germany in all three groups, although the difference is smaller for international relations academics.

Moreover, while we observe statistically discernible differences by field of study ($F(6.00, 7488) = 5.067$, $p < .001$; Rao–Scott adjusted Chi-squared test in the three-category outcome and $F(3, 3744) = 5.9499$, $p < .001$ in binarized, unconditional funding support response variable; Rao–Scott adjusted Chi-squared tests), the general pattern does not vary substantially by field of study. Respondents are slightly more open to U.S.-China collaboration in the field of environmental sciences without conditions other than peer review than collaborations in the other three fields, but the 39.2% of respondents who support these collaborations is markedly lower than the 75.6% who support U.S.-Germany collaborations in the same field.

Figure S4 separates our results in the policy community sample by field of research presented in the vignette. As before, we see remarkably consistent results, with the differences between Germany and China collaborations being both statistically significant and substantively large. We do observe higher levels of support for unconditional funding (in both countries) and a smaller difference between China and Germany in Environmental Science.

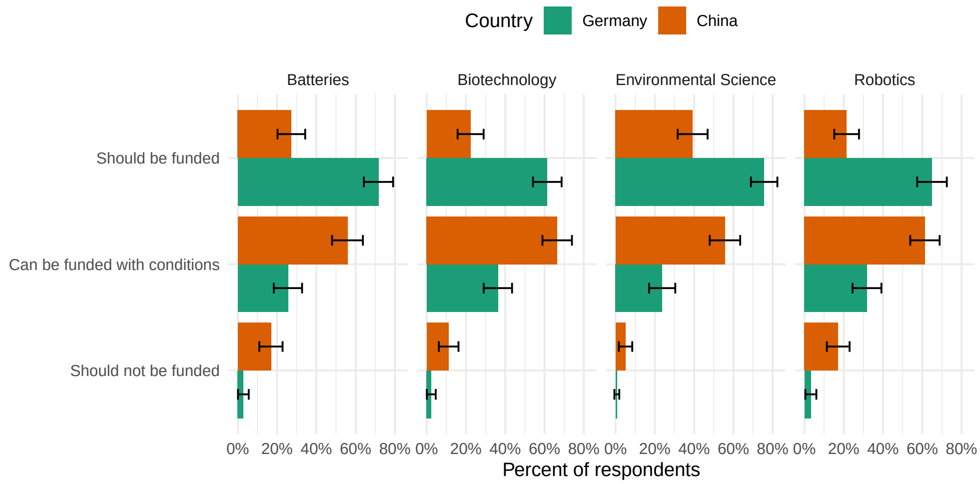


Fig. S4. Policy Community Responses to U.S.-China Collaborations vs. U.S.-Germany Collaborations: By Fields

2.4. Additional Exploration of Heterogeneous Treatment Effects

We conduct an additional exploratory analysis of further heterogeneities in treatment effects based on respondent characteristics.

To explore heterogeneity in respondents' support for funding based on the country of the affiliated scientist, we estimate conditional average treatment effects using both a fully interacted linear model and a machine-learning-based causal forest model. The treatment indicator is a binary variable equal to 1 if the vignette featured a Chinese scientist and 0 otherwise.

We explore heterogeneity across demographic and attitudinal covariates, including gender, age, party identification, respondent group, and both social and economic ideology scales. Observations with missing values on any variables used in the modeling were excluded.

We estimate a causal forest using the `grf` package in R to nonparametrically estimate conditional average treatment effects $\tau(x)$ [3, 4]. The outcome is the binary indicator for whether the respondent believes the scientist "should be funded." Covariates include all demographic, ideological, and professional background characteristics, entered as a sparse design matrix with no intercept. The model is trained with 5,000 trees, using an honesty fraction of 0.8 and a minimum node size of 10.

We assessed calibration using `test_calibration()`, which regresses the target estimand on the mean and differential forest predictions using held-out data. The coefficient on the mean forest prediction was near 1 (0.999, $p < 0.001$), indicating accurate average treatment effect estimation from the causal forest model. However, the non-significant coefficient on differential forest prediction far away from 1 (–0.302, $p = 0.715$) suggests that the model does not well-capture heterogeneity in effects, and serves as an omnibus which does not allow us to reject the null hypothesis of no heterogeneity. The binned calibration plot comparing (Fig. S5) predicted $\tau(x)$ values with doubly robust scores similarly demonstrates that the causal forest model does not capture observed heterogeneity well.

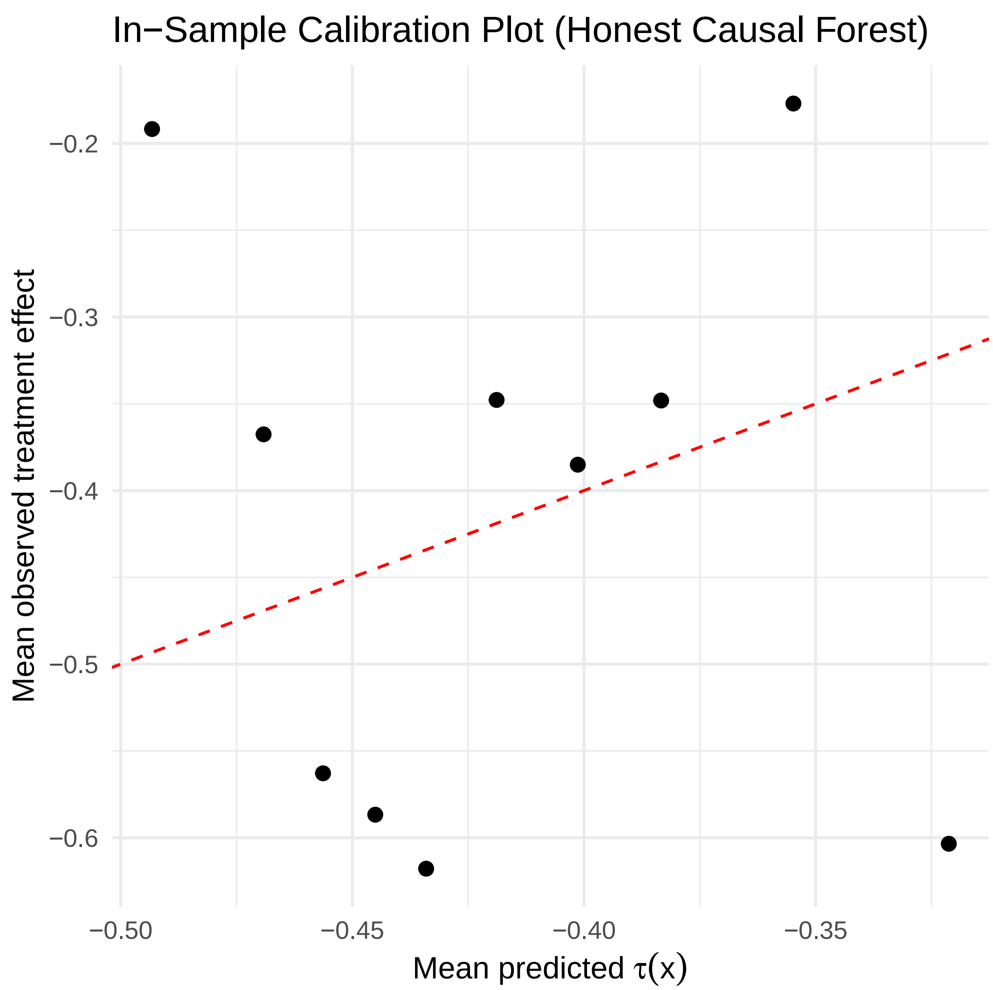


Fig. S5. Binned calibration plot of causal forest model for effect heterogeneity in the policy community

Despite the calibration results not permitting us to reject the null hypothesis of no heterogeneity associated with the covariates in the model, we report the variable importance factors in Fig. S6 below for transparency.

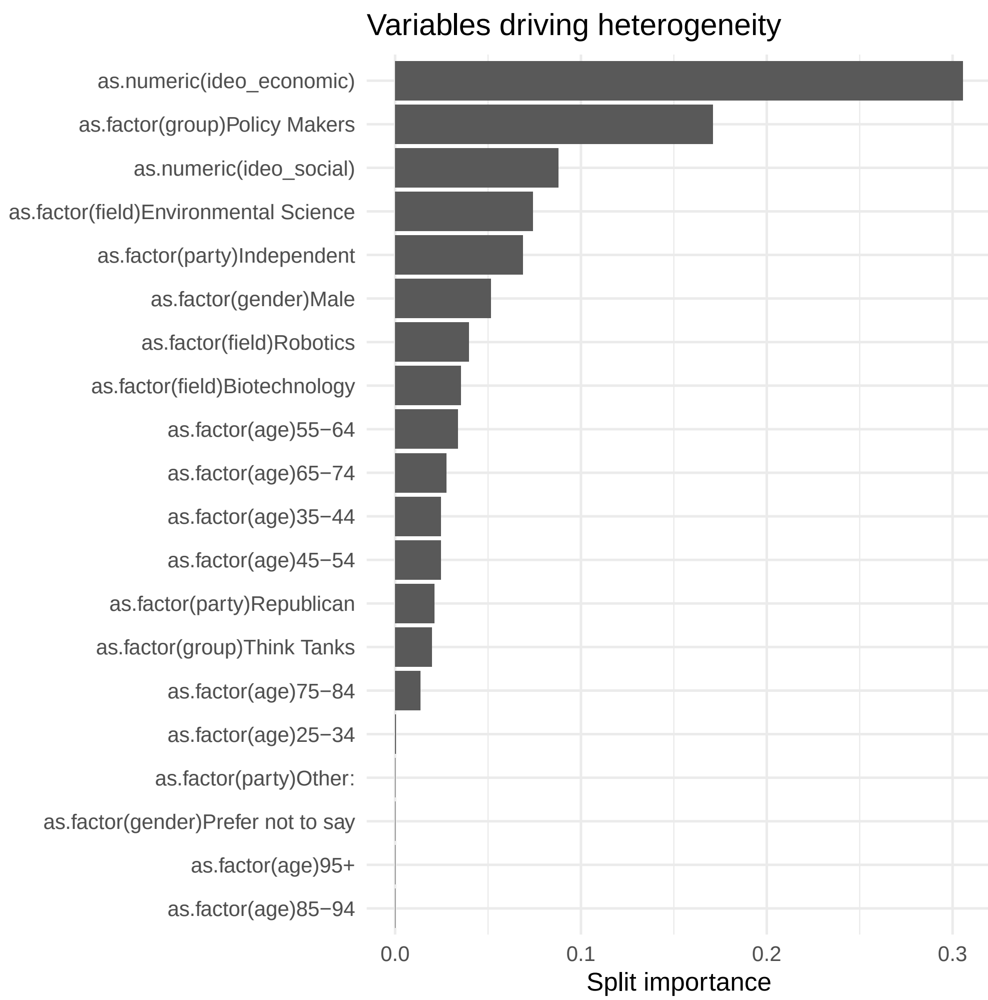


Fig. S6. Variable importance plot from the causal forest model in the policy community

As a robustness check, we also estimate a fully interacted linear regression model including interaction terms between the treatment and all covariates, with results shown in table S1.

While this model imposes stronger parametric assumptions than the causal forest, it yields a substantively similar lack of meaningful heterogenity, with marginally significant differences for Policymakers and "Other" partisans.

3. STUDY 2 - SCIENTIFIC COMMUNITY

The following section reports methodological details of Study 2 as well as supplementary results.

3.1. Experimental Design

The experimental setup in Study 2 was implemented in Qualtrics and was identical to that of Study 1. Figure S7 shows the experimental vignette presented to scientists.

As in Study 1, we randomized two key variables: (1) whether the research proposal involved collaboration between scientists in the U.S. and Germany or between scientists in the U.S. and China, and (2) the scientific field of the collaboration. Respondents were asked under what conditions they believed the proposal should be funded. Respondents who selected the option to conditionally support funding were then prompted to specify any conditions they deemed necessary for the funding to be granted. They could choose from the following options: (1) the collaborator should not have any affiliation with the military or national security establishment, (2) the research should have no potential use in surveillance activities, (3) the research should

	Unconditional Funding Binary
(Intercept)	0.68 (0.12)***
treatment	−0.37 (0.17)*
35-44	0.03 (0.11)
45-54	0.04 (0.11)
55-64	0.04 (0.11)
65-74	0.08 (0.11)
75-84	0.07 (0.13)
85-94	−0.56 (0.46)
95+	0.65 (0.54)
Policy Makers	0.04 (0.06)
Think Tanks	−0.06 (0.06)
Biotechnology	−0.12 (0.05)*
Environmental Science	0.03 (0.06)
Robotics	−0.06 (0.06)
Male	0.13 (0.05)**
Prefer not to say	0.23 (0.17)
Econ_Ideology	0.01 (0.03)
Social_Ideology	−0.05 (0.03)+
Independent	−0.07 (0.05)
Other ×	−0.09 (0.12)
Republican	−0.10 (0.09)
treatment × 35-44	0.05 (0.16)
treatment × 45-54	0.10 (0.16)
treatment × 55-64	0.09 (0.16)
treatment × 65-74	0.13 (0.16)
treatment × 75-84	0.05 (0.19)
treatment × 85-94	0.35 (0.57)
treatment × Policy Makers	−0.17 (0.08)*
treatment × Think Tanks	−0.06 (0.08)
treatment × Biotechnology	0.06 (0.08)
treatment × Environmental Science	0.10 (0.08)
treatment × Robotics	−0.01 (0.08)
treatment × Male	−0.08 (0.07)
treatment × Prefer not to say	−0.39 (0.31)
treatment × Econ_Ideology	−0.07 (0.04)+
treatment × Social_Ideology	0.03 (0.04)
treatment × Independent	0.01 (0.07)
treatment × Other ×	0.28 (0.16)+
treatment × Republican	0.18 (0.13)

Table S1. Fully interacted OLS model of treatment heterogeneity. + p <0.1, * p <0.05, ** p <0.01, *** p <0.001

contribute to a global common good, such as curing cancer or advancing climate science, and (4) the research should be unlikely to be immediately commercialized.

3.2. Sample Characteristics, Weighting, and Balance

To assess representativeness and implement weighting adjustments for the survey responses, we employed the following detailed procedure:

We began by merging our survey data with external data from Web of Science (WoS) on U.S.-based scientists, which we had used as a sampling frame. We then linked the Web of Science data to Dimensions, to add in professional and bibliometric covariates. This involved matching respondents to paper-level WoS identifiers to Dimensions paper identifiers using DOIs as a


Northwestern | Kellogg

In recent years, foreign influence over U.S. science has become an increasing concern for policymakers. Imagine a U.S.-based scientist has proposed a scientific collaboration with scientists in China in the field of robotics to the National Science Foundation. The proposed collaboration does not have an export controlled component and is not classified. Under what conditions do you think the U.S. government should fund such a proposal?

- If the proposal is selected for funding through peer review.
- The research should not be funded, regardless of how well it is reviewed.
- If the proposal is selected for funding through peer review, it should only be funded if the following national security and humans rights conditions are met (click to view and select all that apply)

Fig. S7. Survey for Study 2, First Page

crosswalk. We then used the authors associated with each linked paper in Dimensions to select a pool of candidate researchers in Dimensions, and employed fuzzy string matching via the Jaro-Winkler distance on respondents' names to select the appropriate author on the linked paper for each respondent scientist. The linkage step integrated comprehensive bibliometric and grant-related metadata, including total publications, h-index, grant history, current research organization, and scientific field classification. We are able to link 143,063 of the 218,204 email addresses in our contact list ($\sim$ 65%) to researcher profiles in Dimensions. Of our 7154 respondents we are able to match 73.1 percent to Dimensions researcher IDs.

Survey respondents were categorized based on their completion status, distinguishing between individuals who fully completed, partially completed, or did not respond to the survey. Gender inference was systematically conducted using the `gender_guesser` Python library, applied sequentially to multiple formats of respondents' first names extracted from the dataset (primary listed first names, alternate listed first names, and initials), prioritizing the most confident gender assignments while flagging uncertain or ambiguous cases. The output from the gender inference was transformed into three broad categories for weighting purposes: names classified as "female" or "mostly_female" were categorized as "female"; names classified as "male" or "mostly_male" were categorized as "male"; and names identified as "unknown" or "andy" (ambiguous) were categorized as "unk/andy."

We evaluated sample balance across critical demographic and professional variables, specifically gender (3-category), total publications, total grants, h-index, years of first and last publications/grants, and scientific field. This involved performing chi-squared tests and calculating Cramér's V statistic to quantify associations for categorical variables, which ranges from 0-1, with 0 denoting no association and 1 denoting a perfect association [5]. For continuous variables, we employed independent samples t-tests and calculated standardized mean differences (SMD) to assess the magnitude of imbalance. Balance assessments were conducted both without weights (unweighted) and subsequently with survey weights (weighted).

We show the unweighted balance assessments in Table S2, in which we bin respondents into quartiles (and include a fifth missing category) for scalar variables. It also shows balance using the full range of scalar variables among respondents for whom we have data.

Variable	p-value	χ^2	df	Cramér's V
Gender	1.626×10^{-38}	174.021	2	0.02824
# Pubs Quartile	5.549×10^{-135}	629.782	4	0.05372
# Grants Quartile	1.422×10^{-169}	777.567	2	0.0597
First Pub Year Quartile	8.165×10^{-213}	989.113	4	0.06733
Last Pub Year Quartile	9.176×10^{-46}	207.405	2	0.03083
First Grant Year Quartile	1.191×10^{-228}	1062.183	4	0.06977
Last Grant Year Quartile	1.296×10^{-202}	942.040	4	0.06571
H-index Quartile	2.009×10^{-141}	659.537	4	0.05498
Field (l0)	8.677×10^{-141}	729.258	21	0.07153

Table S2. Unweighted balance results in binned variables

Variable	p-value	$\text{mean}_{\text{resp}}$	$\text{mean}_{\text{nonresp}}$	$\text{mean}_{\text{diff}}$	t-statistic	smd
# Pubs	1.121×10^{-27}	90.943	71.707	19.236	-10.946	0.140
# Grants	6.910×10^{-53}	2.890	1.849	1.042	-15.427	0.205
First Pub Year	2.529×10^{-163}	2005.495	2010.093	-4.598	27.917	-0.365
Last Pub Year	1.754×10^{-46}	2024.120	2023.986	0.134	-14.4084	0.166
First Grant Year	5.871×10^{-17}	2005.058	2007.110	-2.052	8.407	-0.155
Last Grant Year	7.754×10^{-5}	2022.125	2022.612	-0.487	3.956	-0.073
H-index	4.074×10^{-49}	21.701	17.268	4.433	-14.837	0.189

Table S3. Unweighted balance results in scalar variables

Both Tables S2 and S3 demonstrate significant imbalance between respondents and non-respondents along all demographic and scholarly dimensions in our data, although Cramér's V between .028 and .072 in the categorical associations suggest a relatively small substantive imbalance. In the scalar results, we see the largest imbalance in terms of standardized mean difference in first year of publication (2005 for respondents and 2010 for non-respondents), number of grants received (2.89 among respondents and 1.85 among non respondents) and H-index (21.1 among respondents and 17.27 among non-respondents). In general, respondents are older, more productive scholars than non-respondents.

To address this imbalance in our sample, we developed post-stratification weights based on a stratification framework in which respondents were grouped into strata defined by combinations of quartiles for h-index, year of first publication, and scientific field. Each was represented as a categorical variable in these four quartiles, with a fifth category added for missingness. This stratification scheme ensured that the weighted sample of respondents more closely mirrored the distribution of these variables in the target population from our Web of Science contact list. We calculated these weights using the `postStratify` function from the `survey` package in R.

After applying weights, we repeated the balance assessments using weighted chi-squared and weighted t-tests to verify enhanced representativeness and reduce biases introduced by non-response. Weighted results are presented in Tables S4 and S5.

The application of weights dramatically improves balance along all dimensions. In our categorical assessment shown in Figure S4, we no longer observe a statistically distinguishable difference at conventional levels ($\alpha = .05$) in publication count quartile, first publication year, field of research, and H-index. Observed differences are also much smaller, with Cramér's V below .025 in all cases.

Variable	p-value	χ^2	df	Cramér's V
Gender	2.660×10^{-14}	76.608	2	0.019
# Pubs Quartile	0.060	10.910	4	0.007
# Grants Quartile	1.762×10^{-11}	53.673	2	0.016
First Pub Year Quartile	0.848	1.685	4	0.003
Last Pub Year Quartile	0.022	8.967	2	0.006
Last Grant Year Quartile	6.586×10^{-27}	115.978	4	0.023
Last Grant Year Quartile	2.133×10^{-19}	84.938	4	0.020
H-index Quartile	0.914	1.156	4	0.002
Field (l0)	0.999	3.316	21	0.004

Table S4. Weighted balance results in scalar variables

Variable	p-value	$\text{mean}_{\text{resp}}$	$\text{mean}_{\text{nonresp}}$	$\text{mean}_{\text{diff}}$	t-statistic	smd
# Pubs	0.503	72.734	71.707	1.027	0.670	0.002
# Grants	2.003×10^{-5}	2.073	1.849	0.224	4.265	0.013
First Pub Year	3.096×10^{-7}	2009.270	2010.093	-0.822	-5.118	-0.016
Last Pub Year	1.075×10^{-28}	2024.100	2023.986	0.115	11.116	0.034
First Grant Year	0.844	2007.059	2007.110	-0.051	-0.197	-6.005×10^{-4}
Last Grant Year	0.6134	2022.671	2022.612	0.058	0.504	0.002
H-index	0.454	17.473	17.268	0.205	0.749	0.002

Table S5. Weighted balance results in scalar variables

Table S5, examining scalar variables for those for which we have coverage shows similar results. Post-weighting covariates are well balanced between respondents and nonrespondents, with small mean differences and standardized mean differences (SMDs). While the differences between non-respondents for first publication year, last publication year, and number of grants remain significant, their magnitudes are substantially reduced.

3.3. Supplementary Results

We observe a modest but significant association between field of research and experimental funding support in both the categorical (F(6.00, 42,890) = 2.51, p = .020; Rao–Scott adjusted Chi-squared test) and unconditional funding support binary dependent variable (F(3.00, 21,456) = 4.26, p = .005; Rao–Scott adjusted Chi-squared test). However, as with policymakers, these differences are substantively much smaller than the country of collaboration differences, with more support for unconditional funding for U.S.-Germany collaborations than U.S.-China collaborations, and conditional funding more often supported for U.S.-China collaborations regardless of field (see SM Fig. S8). In general, scientists are more supportive of unconditional funding for international collaborations in environmental science, and less supportive of unconditional international collaborations in robotics and biotechnology. However, this difference between fields is consistent for both U.S.-China collaboration and U.S.-Germany collaborations.

As shown in Table S6, there is no significant interaction between the country of the proposed collaboration and the field of research. This null finding is robust to the inclusion of several scientist-level covariates, including primary field of research, year of first publication, h-index, and total number of grants received.

Model 2 indicates that scientists with a greater history of grant acquisition are more likely to support unconditional funding. Specifically, a one standard deviation increase in the number of grants received (6.03 grants) is associated with a 12.2 percentage point increase in the probability of supporting funding for proposals selected through peer review.

Model 3 introduces an interaction between the number of grants received and the country of the proposed collaboration. The results suggest that the positive association between grant experience and support for unconditional funding is primarily driven by respondents evaluating proposals involving German collaborators.

Models 4-6 repeat these models without post-stratification weights, demonstrating that our results are consistent whether or not we use the weights.

Figure S8 breaks out our three category response results by field presented in the vignette, showing a smaller difference between evaluations for China and Germany in collaborations for environmental science research than for other fields.

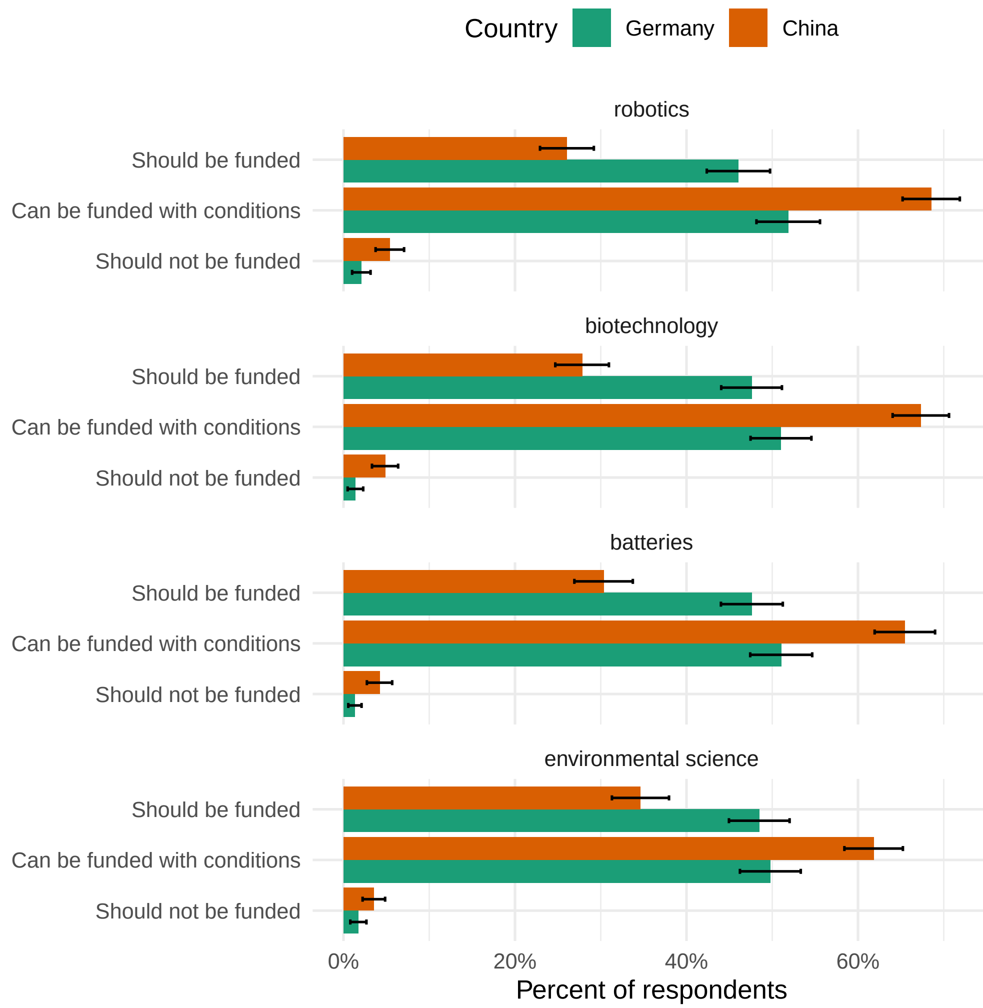


Fig. S8. U.S.-China Collaborations vs. U.S.-Germany Collaborations: By Fields. Error bars indicate bootstrap 95% CIs.

3.4. Additional Exploration of Heterogeneous Treatment Effects

We explore heterogeneity in respondents' support for funding scientists based on the country affiliation of the scientist featured in the vignette, using causal forest models from the `grf` package in R. The binary treatment variable is set to 1 for respondents assigned to a vignette featuring a Chinese scientist and 0 otherwise. The binary outcome is whether the respondent unconditionally supports funding the featured scientist.

The modeling sample includes demographic (age, gender, race/ethnicity), ideological (economic and social ideology, 7-category liberal to conservative scales), three-category party iden-

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	−0.832***			−0.777***		
	(0.082)			(0.057)		
Germany	0.736***	0.729**	0.649**	0.810***	0.803***	0.713***
	(0.111)	(0.227)	(0.231)	(0.049)	(0.140)	(0.145)
biotechnology	−0.122	−0.255+	−0.252+	−0.108	−0.302***	−0.301***
	(0.114)	(0.137)	(0.138)	(0.070)	(0.086)	(0.086)
environmental science	0.196+	0.067	0.075	0.103	0.081	0.086
	(0.111)	(0.166)	(0.164)	(0.069)	(0.098)	(0.097)
robotics	−0.211+	−0.377*	−0.375*	−0.170*	−0.377***	−0.377***
	(0.117)	(0.165)	(0.165)	(0.071)	(0.111)	(0.112)
Germany × biotechnology	0.120	0.142	0.140		0.159	0.159
	(0.154)	(0.212)	(0.211)		(0.141)	(0.140)
Germany × environmental science	−0.162	−0.104	−0.102		−0.099	−0.093
	(0.152)	(0.326)	(0.323)		(0.210)	(0.208)
Germany × robotics	0.148	0.165	0.167		0.146	0.150
	(0.158)	(0.328)	(0.328)		(0.217)	(0.217)
First Pub Year		−0.005+	−0.005+		−0.004	−0.004
		(0.003)	(0.003)		(0.003)	(0.003)
H-index		−0.001	−0.001		−0.002	−0.002
		(0.002)	(0.002)		(0.002)	(0.002)
# Grants		0.019**	0.003		0.020**	0.007
		(0.007)	(0.010)		(0.007)	(0.009)
CountryGermany × # Grants			0.030***			0.025***
			(0.005)			(0.006)
Num.Obs.	7154	5212	5212	7154	5212	5212
R2	0.027	0.007	0.008	0.031	0.065	0.065
R2 Adj.	0.026	−0.004	−0.003	0.030	0.056	0.056
R2 Within		0.036	0.037		0.038	0.039
R2 Within Adj.		0.032	0.033		0.035	0.036
AIC	9289.3	5994.2	5989.4	9350.0	6629.8	6625.4
BIC	9352.4	6204.0	6205.8	9395.3	6839.7	6841.8
Log.Lik.	−4636.272			−4671.012		
F	29.913			72.061		
RMSE	0.48	0.47	0.47	0.48	0.47	0.47
Std.Errors		by: field	by: field		by: field	by: field
FE: field		X	X		X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S6. Logistic regression results of binarized DV corresponding to whether respondent selected that a proposal should be funded 'If the proposal is selected for funding through peer review.'

tification, and professional characteristics (field, publication and funding history, h-index, year of first publication) of the respondent, along with indicators of prior international collaboration. These career variables are retrieved from Dimensions for the set of respondents we were able to match to Dimensions researcher IDs. We discretized prior collaboration into binary indicators (0 = no collaboration; 1 = any collaboration) for collaboration with non-U.S., Chinese, and German coauthors. Racial and ethnic identity variables were recoded into indicator variables for White, Black, Asian, Latino, Other, and Declined to Answer. Publication counts, h-index, and total number of grants were log transformed prior to modeling. Observations with missing values were removed prior to analysis.

We estimate a causal forest with 5,000 trees, using respondent-specific sampling weights as provided in the survey design. The model includes a rich set of covariates capturing demographic, ideological, and professional characteristics. Covariates were entered via a sparse design matrix without an intercept.

Predicted individual-level conditional average treatment effects (CATEs), denoted $\tau(x)$, were extracted using in-sample predictions. Doubly robust scores were also computed for use in calibration assessment.

Model calibration was evaluated using `test_calibration()`, which regresses observed treatment effects on both the mean and differential forest predictions in a held-out sample. The coefficient on the mean prediction was near 1 (1.017, $p < 0.001$), indicating accurate estimation of the average treatment effect, while the significant and large coefficient on the differential prediction (0.871, $p < 0.001$) suggests that the model captures meaningful heterogeneity in treatment effects across respondents. Together, these results suggest the existence of systematic heterogeneity detectable at the omnibus level.

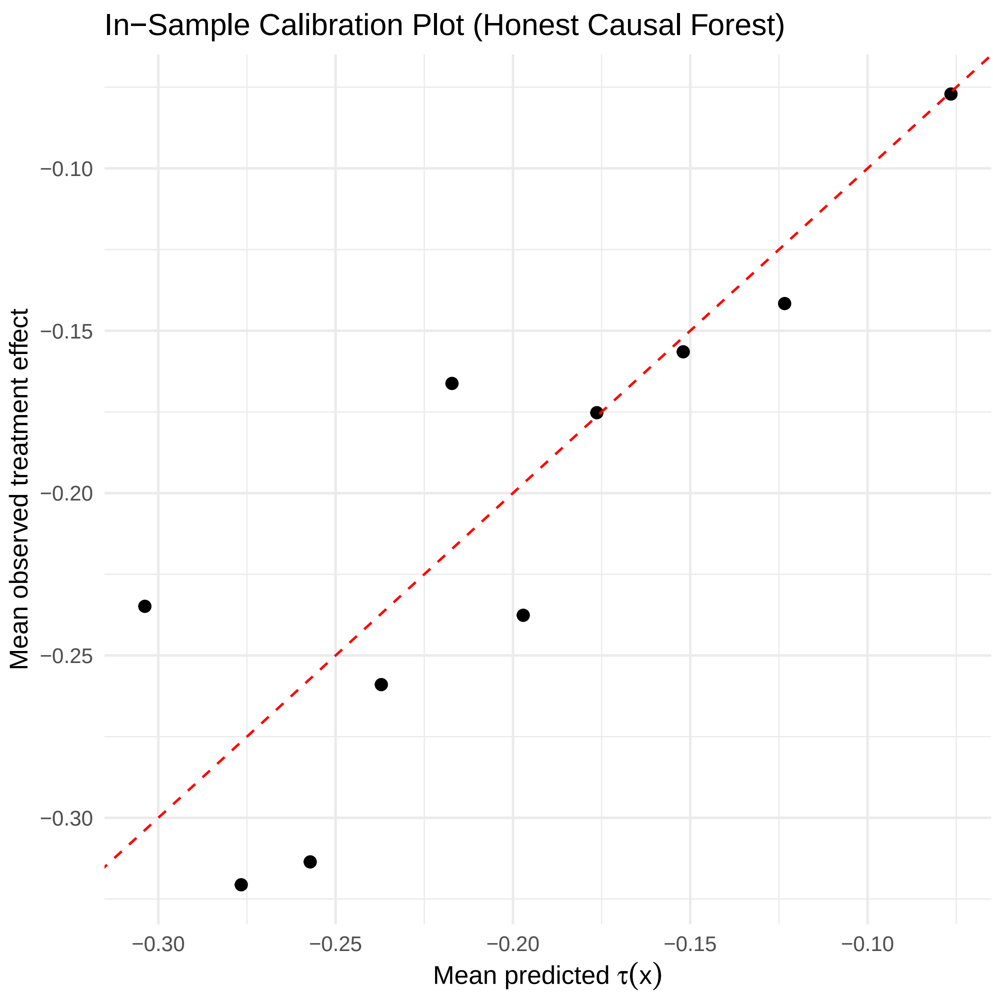


Fig. S9. Binned calibration plot of causal forest model for effect heterogeneity in the scientific community

A binned calibration plot confirms overall alignment between predicted and observed treatment effects (Figure S9).

We assessed covariates contributing most to heterogeneity using forest split importance scores (Figure S10). The top drivers included ideological orientation, collaboration history, and field of study.

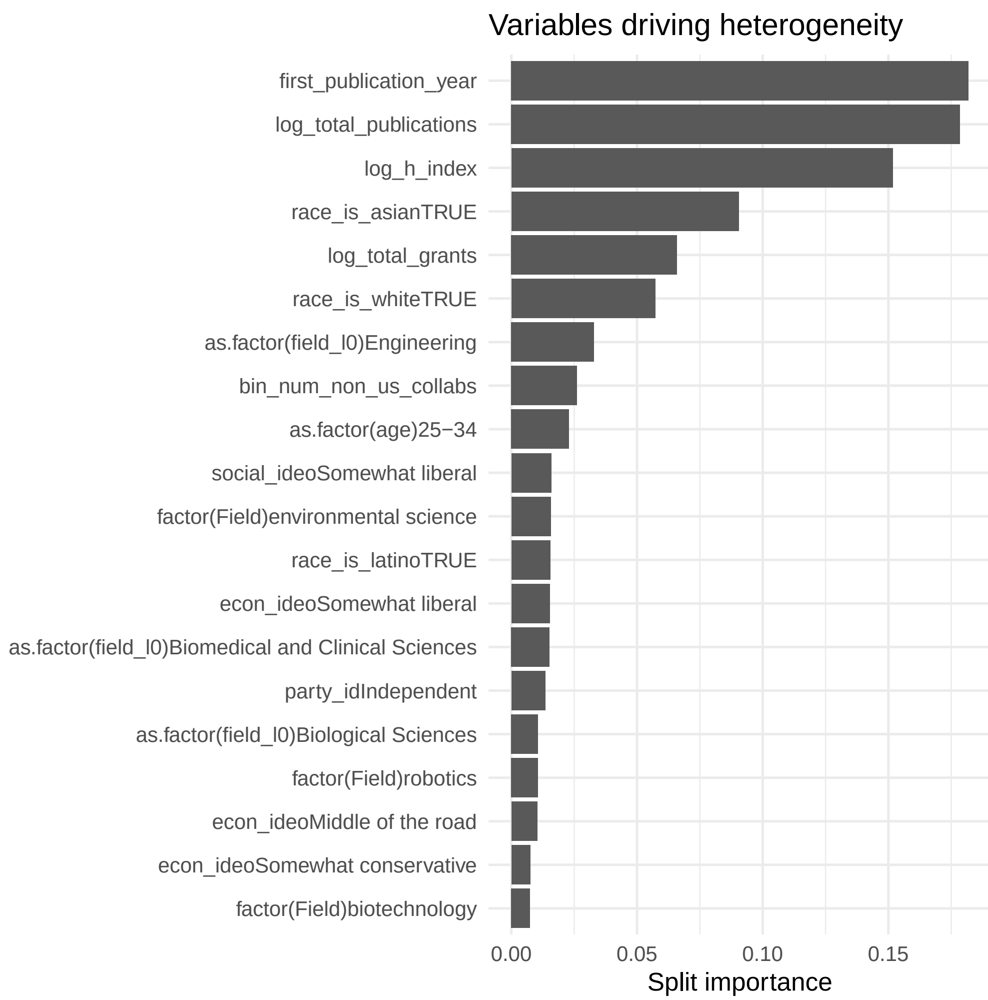


Fig. S10. Variable importance plot from the causal forest model in the policy community

We visualized heterogeneity in $\tau(x)$ across key numeric and categorical covariates. For continuous characteristics—such as number of publications, h-index, grant count, and seniority—treatment effects were smoothed using LOESS (Figure S11).

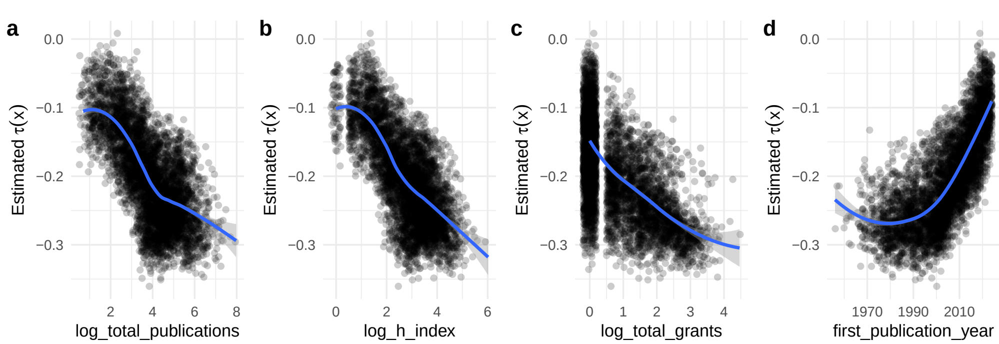


Fig. S11. Variable importance plot from the causal forest model in the policy community

For categorical variables (e.g., ideology, party ID, collaboration history, race/ethnicity), point intervals summarize conditional treatment effects across subgroups (Figure S12).

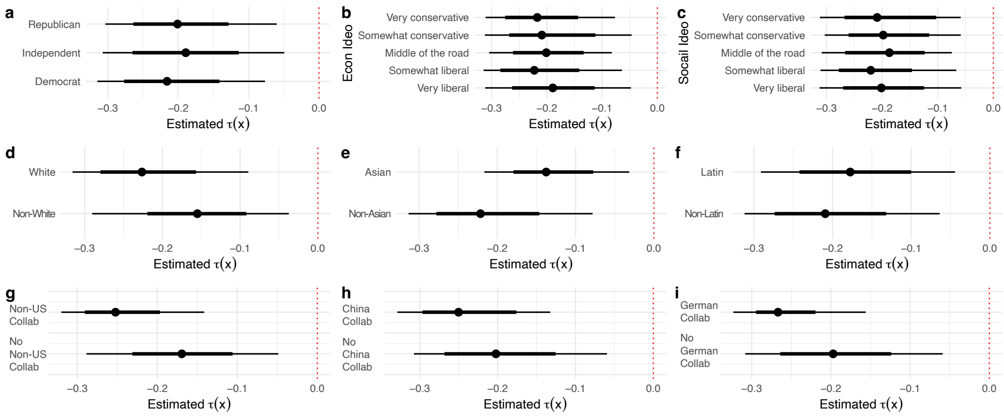


Fig. S12. Variable importance plot from the causal forest model in the policy community

Additional subgroup-specific $\tau(x)$ estimates are displayed by field of study, vignette field, and age (Figure S12). Fields are ordered by estimated $\tau(x)$, and plots include 95% uncertainty intervals.

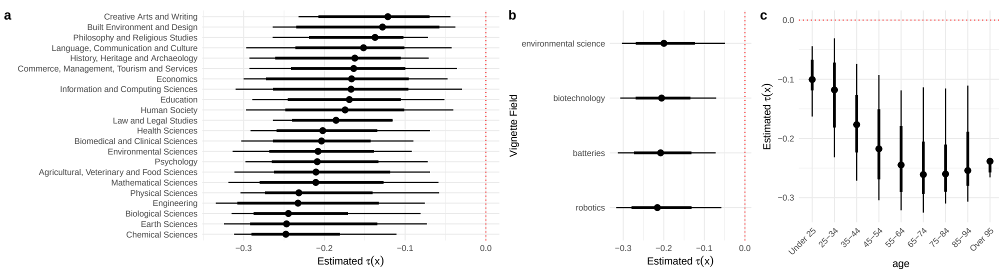


Fig. S13. Variable importance plot from the causal forest model in the policy community

Together, these results offer descriptive insight into how respondent characteristics shape their willingness to fund scientists from China, while there is some statistically significant heterogeneity. The effect tends to be larger among older scholars, with a longer history of publication, more grants and a higher h-index. Additionally, on average the effect is larger among white scholars and smaller among Asian scholars. However, the estimated conditional average treatment effect is always negative and large in all groups.

Finally, we estimate a fully interacted model with field fixed effects on the unconditional funding support outcome, which we present in Table S7 below. For space we display only the interaction terms.

REFERENCES

1. P. C. Avey, I. E. G. Balnes, S. N. Gillooly, *et al.*, "Trip policymaker survey report," Teaching, Res. Int. Policy (TRIP) Proj. Glob. Res. INstitute. Available at https://trip.wm.edu. (2023).
2. P. C. Avey, M. C. Desch, E. Parajon, *et al.*, "Does social science inform foreign policy? evidence from a survey of us national security, trade, and development officials," Int. Stud. Q. **66** (2022).
3. S. Athey and S. Wager, "Estimating treatment effects with causal forests: An application," Obs. studies **5**, 37–51 (2019).
4. S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," J. Am. Stat. Assoc. **113**, 1228–1242 (2018).
5. H. Cramér, *Mathematical methods of statistics*, vol. 9 (Princeton university press, 1999).

	DV: binary unconditional funding support
treatment × 25-34	−2.222 (1.016)*
treatment × 35-44	−2.531 (0.978)**
treatment × 45-54	−2.331 (0.988)*
treatment × 55-64	−2.396 (0.904)**
treatment × 65-74	−1.970 (1.004)*
treatment × 75-84	−2.009 (1.014)*
treatment × 85-94	−2.230 (1.329)+
treatment × Over 95	−25.978 (0.984)***
treatment × Independent	0.011 (0.131)
treatment × Republican	−0.488 (0.476)
treatment × EconIdeo-Somewhat liberal	−0.062 (0.264)
treatment × EconIdeo-Middle of the road	0.148 (0.238)
treatment × EconIdeo-Somewhat conservative	0.467 (0.314)
treatment × EconIdeo-Very conservative	0.636 (0.650)
treatment × EconIdeo-Somewhat liberal	−0.437 (0.174)*
treatment × SocialIdeo-Middle of the road	−0.466 (0.257)+
treatment × SocialIdeo-omewhat conservative	−0.523 (0.274)+
treatment × SocialIdeo-Very conservative	−1.175 (0.507)*
treatment × batteries	0.291 (0.438)
treatment × biotechnology	0.120 (0.278)
treatment × environmental science	0.445 (0.214)*
treatment × log(#publications)	0.042 (0.070)
treatment × log(h-index)	−0.079 (0.174)
treatment × log(#grants)	−0.047 (0.081)
treatment × First Publication Year	0.013 (0.015)
treatment × race_is_white	−0.346 (0.271)
treatment × race_is_black	−0.142 (0.405)
treatment × race_is_latino	0.282 (0.289)
treatment × Non US Collaboration	−0.075 (0.178)
treatment × China Collaboration	0.125 (0.149)
treatment × Germany Collaboration	−0.106 (0.213)
treatment × race_is_asian	0.435 (0.298)
treatment × race_is_other	−0.187 (0.323)
treatment × race_is_decline	−0.664 (0.416)
Num.Obs.	4746
R2	0.043
R2 Adj.	0.010
R2 Within	0.070
R2 Within Adj.	0.045
AIC	5342.2
BIC	5930.5
RMSE	0.46
Std.Errors	by: field_l0
FE: field_l0	X

Table S7. Fully interacted OLS model of treatment heterogeneity. Only interaction terms displayed for space. + p <0.1, * p <0.05, ** p <0.01, *** p <0.001