For Francesca, my compass.

Table of Contents

Acknowledgments

Writing this dissertation has been as much a personal journey as an academic one. To some extent these words, mark the culmination of years of work, learning, growth, and many moments of both uncertainty and breakthrough. None of this would have been possible without the support, guidance, inspiration, and encouragement of many people. While it is impossible to name everyone who has been a part of this process, I am deeply grateful to all those who played a role, big or small, in this journey.

I am especially grateful to my supervisor, Simon Munzert, for his unconditional support, generosity, and steady mentorship. I could not have asked for a better advisor, mentor, collaborator, and role model—someone who trusted my independence, reassured me in moments of doubt, and offered thoughtful guidance when it mattered most. His patience, encouragement, and belief in my potential helped me stay grounded and grow into this work. For all of this, and more, I am deeply thankful.

Throughout these years, my family has been a constant source of motivation and strength. Though separated by distance, I have always felt their presence, and they have remained an integral part of this journey. I am deeply grateful to Mami, Papi, Nana, Nano, Moni, Nico, Rena, and Hugo for their unwavering support, for nurturing my curiosity, and for shaping me into the person I am today. I can never fully repay their patience during the long stretches apart and the many moments we could not share in person. Their steady belief in me, even from afar, made all the difference. I am also thankful to my family in the Bisiacaria for giving me a home away from home.

I have been incredibly fortunate to be part of a vibrant and generous academic community. My colleagues, collaborators, advisors, and the members of Data Science Lab enriched this work through thoughtful conversations, co-authorship, and critical exchange. I am also grateful to the PhD team for their tireless support behind the scenes, which made the path smoother. Also important, my 'Berlin family' who reminded me—often over food, laughter, long walks, bike rides, and, in the truest sign of my full Brandenburg integration, mushroom picking—that there is more to life than research. Their friendship grounded me and made these years not just productive, but truly happy.

Perhaps the most important companion in this entire journey is my favorite person, best friend, advisor, editor, and collaborator—who also happens to be my wife—Francesca. We embarked on this path together, not knowing of what was ahead. Her constant support and brilliant mind have been my compass throughout this process, always guiding me forward when the way was unclear. Francesca is bright, strong, thoughtful, impressive, loving, and selfless in ways I do not think she fully realizes. She makes everything better simply by being herself. This dissertation could never have come to life without her—and I am 100% sure her unexplainable ability to find four-leaf clovers had something to do with it, too.

Executive Summary

While evidence can be a critical input for policymaking, our understanding of how it flows from research and expert communities to elite decision-makers remains remarkably limited. Studying these flows, and the relationships that underpin them, is inherently difficult. They are fragmented, politically mediated, and largely unobservable. This dissertation addresses these challenges by advancing a data-extensive computational social science approach, drawing on alternative (largely unstructured) data sources to systematically observe and characterize elite–evidence interactions. This project consists of four studies, involves the creation of new data infrastructures, and provides a foundation for future research and a proof of concept for investigating these dynamics. Collectively, these studies offer new perspectives on who gets seen and heard, what kinds of expertise gain traction, and how institutional and individual factors can shape evidence use in practice.

In the first paper, I analyze behavioral data from nearly 3,700 legislators across 12 countries to examine how elected officials follow and interact with academic researchers on Twitter. These data offer a scalable and less obtrusive lens for observing elite–researcher connections that are typically hidden from view. The findings reveal that while many legislators follow researchers, visible engagement is rare—and strongly related to their partisanship and educational background. Legislators from conservative and radical-right parties are significantly less likely to follow or interact with scholars online. Importantly, I use the outset of the COVID-19 pandemic as an exogenous shock to information demand, showing that elite attention to medical expertise increased measurably during this period. This paper illustrates how informal and digitally mediated channels can serve as windows into elite information-exposure and engagement behavior, offering a richer understanding of how (and when) evidence producers enter the periphery of political attention.

In the second paper, Roman Senninger and I extend the analytical focus from individual behavior to institutional outputs. We map the evidence base cited in 1.2 million policy documents across 185 countries. Using large-scale citation and institutional data, we document a striking concentration in referencing practices: both academic and policy sources referenced in these documents are overwhelmingly produced in the global North. Even governments in the global South predominantly cite Northern research. While this is not a new claim in global knowledge politics, the scope and granularity of our dataset allow us to systematically trace how epistemic asymmetries are codified in the very documents that inform policy action.

In the third paper, I examine a more formalized setting of evidence engagement: parliamentary expert hearings. I introduce a novel database—the Bundestag Expert Witness Tracker (BEWIT)—which captures all public expert hearings in the German Bundestag from 2009 to 2024. These institutional data provides a unique view into who is called upon to present knowledge in the legislative process. I illustrate how political gatekeeping and transparency reforms shape the informational ecosystem of the German parliament by linking experts to their academic profiles, lobbying register, and the political groups that bring them forward. The analysis not only speaks to long-standing theories of legislative information-seeking but also showcases how data infrastructures can make previously opaque practices traceable.

Finally, the fourth paper shifts from observing elite behavior to measuring individuals' capacity to engage with evidence. In this collaborative endeavor, Simon Munzert and I develop and validate the INSPIRE instrument—a psychometrically robust measure of scientific evidence consumption literacy tailored to public policy contexts. We test the tool across general population samples and policy-adjacent professional groups, providing both a diagnostic and an applied framework for understanding evidence reasoning in complex decision-making environments.

Together, these studies contribute to a growing literature that studies the interface between knowledge and policy decision-making. They reveal not only how policymakers encounter evidence, but also how structural inequalities, institutional practices, and individual characteristics and capacities shape those interactions. This empirical portfolio supports the dissertation's broader aim of showing how elites engage with evidence (and its providers), tracing both the observable micro-behaviors of individuals and the systemic forces that mediate knowledge flows in contemporary governance.

Preface

Evidence can be a critical input in policymaking. Whether in conceiving climate legislation, regulating food standards, or responding to public health crises, policymakers frequently invoke the authority of scientific knowledge and technical expertise (Cairney, 2016). However, despite its rational appeal, the notion of policymaking guided by evidence often obscures more than it reveals. What counts as evidence? Whose expertise permeates decisions? How does knowledge flow—if at all—from research and expert communities to elite decision-makers? And do policymakers have the competencies to work with evidence in the first place? The answers to these questions are far from straightforward. In practice, the flow of evidence into policy is fragmented, contested, and deeply shaped by larger societal, political, institutional, and cognitive dynamics (Caplan, 1979; Weiss, 1979; Head, 2008; Contandriopoulos et al., 2010; National Research Council, 2012; Parkhurst, 2017).

Across a wide range of policy theories, information is seen as more than just a factual input—it helps structure how issues are framed, which alternatives are considered, and whose voices carry weight in decision-making (e.g., Simon, 1944; Lindblom, 1959; Jones and Baumgartner, 2005; Cairney, 2012). Elite policy decision-makers routinely face decisions that require engagement with specialized information (Baumgartner and Jones, 2015). The reality that such actors operate in evidence-rich environments interacts with further tensions. The use of evidence is not just a matter of access or supply—it is also a matter of politics, structure, and individual capacity. Even as democratic systems adopt formal commitments to evidence-informed decision-making (e.g., Juncker, 2014; Kenny et al., 2017; U.S. Congress, 2019), we still know remarkably little about how information actually reaches, influences, or is filtered by those at the heart of public power.

These gaps are particularly salient in efforts to understand the interactions between expert knowledge and policy elites. Studying these elite-evidence interactions presents both conceptual and methodological challenges. Much of the relevant dynamics often unfold behind closed doors or are embedded in institutional routines and symbolic practices that are difficult to observe. Established research methods, including both quantitative and qualitative approaches such as interviews, surveys, document analyses, and more recently, survey and field experiments, have yielded important findings (e.g., Bogenschneider, Day, and Bogenschneider, 2021; Hjort et al., 2021; Geddes, 2023; Vivalt and Coville, 2023; Ban, Park, and You, 2024; Furnas, LaPira, and Wang, 2025). The findings from research employing these approaches provide depth, context, and access to the perspectives of key actors. Yet they face important limitations: they often rely on retrospective, or self-reported, accounts, are difficult to scale across countries or periods, and struggle to capture the fluid, real-time dynamics through which information flows in complex political systems.

This dissertation responds to these challenges by advancing a data-extensive approach to studying how evidence permeates the policymaking process through elite actors. Motivated by the practical and analytical limitations of existing strategies, I draw on new data sources —such as digital trace data, policy metadata, and unstructured text—and apply computational methods to systematically observe and characterize elite–evidence interactions at scale. Rather than relying on retrospective interviews or self-reports, this approach enables direct observation of the settings in which decision-makers come into contact with expert information, and with whose expertise they engage in practice. Crucially, this work also involves the creation of novel data infrastructures, drawing on extensive sources that are dispersed, unstructured, and not readily usable in their original form. These carefully assembled data sources make a significant contribution in their own right, offering both a foundation for future research and a proof of concept for studying elite–evidence dynamics at scale.

To make sense of the scope of my dissertation, it helps to clarify the actors and types of evidence at its center. The studies focus on two groups of elite policy decision-makers: legislators and policy professionals. While most of the observable interactions in the papers of this dissertation involve elected officials, the studies also trace signals from tools, platforms, and institutional practices developed by—and aimed at—the broader policy workforce: behind-the-scenes staffers, analysts, and advisors who help shape the technical and strategic contours of legislation and policy. As key players through which evidence comes about and enters political systems, the practices and priorities of these two groups have been widely studied as central to the uptake of research and expertise in policy settings (see Oliver et al., 2014; Ouimet et al., 2023). Looking into the spaces where legislators and policy professionals interact with expert knowledge is, therefore, essential for unpacking the ways through which evidence makes its way into the policy arena.

Equally important is what this project defines as "evidence". While evidence can take many forms, my project's primary focus is on the scholarly end—that is, academic research and the researchers who produce it. This emphasis is partly personal: my doctoral journey began at the onset of a global pandemic, a moment that called for deep reflection on the public role of expertise. I approached these questions as a quantitative social scientist working within a public policy school. This environment exposed me daily to the practical challenges of applying evidence to real-world problems. Alongside my research, I spent a significant amount of time teaching causal inference and data science to policy professionals from around the world. That experience led me to think critically about what it means to produce knowledge intended to serve the public good. That said, embedded in these projects is the recognition that decision-makers engage with a broader ecosystem of expertise where academic research is one input among many (Weiss, 1977; Head, 2008; Baumgartner and Jones, 2015). In addition to scholarly sources, the studies also integrate and examine other forms of evidence, such as knowledge produced by governments themselves, interest groups, think tanks, nongovernmental organizations, and intergovernmental bodies. Across the papers, I examine the diversity of evidence sourcing, while highlighting the analytical and political tensions that arise when different forms of expertise compete for attention in elite decision-making spaces.

Idealized models often depict a rational, knowledge-driven process, in which high-quality research is available and used to guide decisions. However, in practice, the informational environments of decision-makers are shaped by self-selection, institutional constraints, cognitive limits on attention, and a balancing act between policy and political consequences (Gilardi, 2010; Walgrave and Dejaeghere, 2017; Furnas, LaPira, and Wang, 2025). These dynamics determine not only whether evidence is used, but also which knowledge is seen, by whom, and under what conditions. This dissertation opens new analytical ground by focusing on the real-world settings and vehicles—digital spaces, policy briefs, hearings—through which elites encounter, select, and engage with expert knowledge (and experts themselves).

At the core of this project is a new perspective on the ways we can study the evidence–policy interface: from assuming evidence reaches decision-makers, to zooming into settings in which they are likely to encounter and select it. All this, while employing methodological and computational tools that allow us to peek in on these moments in ways that were previously unfeasible. I combine novel data and measurement strategies to examine the informational foundations of policy and political decision-making. The studies in this dissertation offer new insights into who is seen and heard, what kinds of expertise gain traction, and how institutional and individual factors can shape evidence use in practice. Rather than treating "evidence-based policy" as an aspirational norm, the dissertation takes it as an empirical 'puzzle'—one that requires close attention to the messy, contingent, and uneven ways that knowledge travels in policy contexts.

This doctoral project contains four individual papers. Three are empirical studies involving large-scale data collection efforts that merge diverse data sources to uncover signals from different environments where policy decision-makers encounter evidence (and its providers). The fourth is a conceptual and measurement-focused piece, aimed at developing an instrument to assess competencies in evidence consumption in policy contexts. Together, these papers vary along several dimensions: in geographic scope (from the global to the national level), in the granularity of the data (from individual-level interactions to systemic patterns), and in the traceability of observed behaviors. Yet as a collective body of work, they offer complementary indications into overlapping findings, such as how political orientation relates to the willingness to engage with different kinds of expertise, or the relative salience of social science and scientists relative to other disciplines across settings in the studies. In doing so, this dissertation contributes to a growing body of research on the politics of expertise, the flow of expert information in policy environments, and the broader characteristics of the evidence–policy interface.

Evidence in policymaking

In recent decades, the relationship between science, expertise, and policymaking has become central to both scholarly inquiry and political discourse. Across many democratic systems, the latter half of the twentieth century saw a growing emphasis on technical and expert knowledge as a means to modernize statecraft and improve decision-making (Ham, Hunter, and Robinson, 1995; Sanderson, 2002). These rationalist ideals converged around a key notion: that public policy can be made more effective, efficient, and accountable through the systematic use of empirical evidence.

This ideal crystallized in the evidence-based policymaking (EBPM) paradigm, which gained traction across policy systems in the 1990s and 2000s (Davies, Nutley, and Smith, 1999). Rooted in the earlier success of evidence-based medicine, EBPM proposes that government decisions should be grounded in objective research rather than ideology or intuition. It assumes that evidence is readily available, interpretable, and usable—and that policymakers can, in principle, integrate it impartially into the policy process, and evaluate outcomes using scientific methods (Baron, 2018; Bowers and Testa, 2019).

Yet the appeal of EBPM often obscures the complexity and contestation inherent to real-world policymaking. As scholars have long emphasized, evidence does not speak for itself: it must be selected, interpreted, and mobilized within institutional and political environments that condition both its visibility and its influence (Cairney, 2016). In this view, policymaking is shaped by bounded rationality (Tversky and Kahneman, 1974; Janis and Mann, 1977)—a condition under which decision-makers operate with limited information, finite attention, and competing priorities. Rather than objectively weighing evidence, policymakers are often navigating a landscape where ideological commitments, public opinion, electoral incentives, and cognitive biases constrain how evidence is processed and applied (Kraft, Lodge, and Taber, 2015; Bisgaard and Slothuus, 2018; Senninger and Seeberg, 2024). In most contexts, evidence is not a neutral input but one among many factors informing policy choices, frequently competing with (and sometimes subordinated to) political and strategic calculations.

This recognition has led to a rethinking of the EBPM ideal. Critics argue that while the pursuit of evidence-based policymaking is well-intentioned, it can sometimes obscure the inherently political nature of policy decisions (Nevo and Slonim-Nevo, 2011). Others raise concerns about the tension between technocratic governance and democratic accountability (Caramani, 2017), especially in contexts where technical authority risks sidelining public deliberation. Relatedly, scholars have pointed to the increasing politicization of science and the erosion of public trust in expert institutions (Tollefson, 2020; Algan et al., 2021; Cologna et al., 2025). As a response, the concept of evidence-informed policymaking (EIPM) has gained traction. EIPM acknowledges that while evidence is crucial, it must be interpreted, contextualized, and balanced with values, political constraints, and local knowledge (Head, 2016; Bowers and Testa, 2019; Mair et al., 2019).

The social sciences have responded to these tensions with a growing empirical literature on how policymakers respond to evidence in practice. A substantial strand of this work comes from experimental research—both field and survey-based—that seeks to identify when, how, and under what conditions decision-makers update their beliefs in response to new information. In field experiments, researchers have evaluated interventions aimed at bringing policymakers and researchers closer together. For example, Crowley et al. (2021) showed that U.S. Congressional offices randomly assigned to a Research-to-Policy Collaboration program became more likely to engage with science and incorporate evidence in legislation. Similarly, Mehmood, Naseer, and Chen (2021) found that Pakistani civil servants who received training in causal inference scored better on exams, expressed greater willingness to pay for research evidence, and were more likely to choose evidence-backed policies. In Brazil, Hjort et al. (2021) found that informing mayors about RCT evidence on tax compliance increased the likelihood that such policies would be implemented.

Survey experiments offer a broader empirical base and more flexibility in outcome measurement. These studies often use conjoint or discrete choice designs to assess which attributes of evidence matter to policymakers. Findings suggest that while researchers prioritize internal validity and methodological rigor, policymakers tend to value contextual relevance, endorsement by local experts, and applicability to their own settings. Larger sample sizes and local relevance consistently increase the appeal of research findings among policy professionals (Lee, 2022; Vivalt, Coville, and Kc, 2025).

Beyond evidence attributes, another set of studies investigates how policymakers update their beliefs. Here, the results are mixed. Some studies highlight bias and motivated reasoning. For instance, Baekgaard et al. (2019) and Banuri, Dercon, and Gauri (2019) find that even experts are susceptible to confirmation bias, framing effects, and ideological filtering. Others take a more optimistic view, suggesting that policymakers can and do revise their beliefs in light of new evidence, even when it challenges their priors—though the strength and persistence of these effects may depend on how evidence is presented (Nakajima, 2021; Lee, 2022; Vivalt and Coville, 2023).

Taken together, this experimental literature paints a more nuanced picture of evidence use than the EBPM model suggests. It underscores that the uptake of evidence is conditional, uneven, and shaped by institutional, cognitive, and political dynamics. However, it also reveals important limitations.

First, the scope of existing studies is narrow—most focus on a few countries, levels of government, or issue areas. Second, some designs assume high data literacy among policy professionals, testing preferences over technical attributes (e.g., RCT vs. quasi-experimental designs) that may not be clear to many real-world decision-makers. Third, a key motivation for this dissertation, these studies typically assume that decision-makers have already been exposed to research evidence.

This last point represents a critical conceptual and empirical gap. By studying responses to evidence that is directly presented, much of the existing literature estimates the effects of evidence conditional on exposure. Yet in practice, policymakers often self-select into what information they consume, based on incentives, capacities, and information environments (Walgrave and Dejaeghere, 2017). As such, the average treatment effects identified in these experiments tell us little about the dynamics of real-world evidence engagement—the messy, unstructured processes by which policymakers encounter (or fail to encounter) expert information 'in the wild'.

This concern is a central motivation for the present dissertation. Rather than concentrating solely on belief updating or retrospective evaluations of policy outcomes, I point the analytical lens to an earlier, often overlooked stage: how, when, and under what conditions elite decision-makers encounter, are exposed to, or actively engage with evidence and its providers. This approach moves beyond stylized accounts that treat evidence as a static input by foregrounding encounters and exposure. This shift reframes the evidence–policy relationship as not just a matter of use, but also visibility, access, and self-selection.

Indeed, studies suggest that familiarity with and trust in the information source are key drivers of evidence uptake (Gollust et al., 2017; Crowley et al., 2021). The accessibility of knowledge in institutional environments—such as parliamentary hearings or committee briefings—also shapes opportunities for engagement (Kenny et al., 2017; Geddes, 2018; Geddes, 2023). Increasingly, digital spaces play an important role in how evidence is circulated and encountered across the globe (Mede et al., 2024).

Studying the interaction between evidence and policymaking poses distinct methodological challenges. Policymakers are a hard-to-reach population, and conventional tools—like survey self-reports—face well-known limitations, including social desirability and recall bias. For example, while surveys asking MPs how often they rely on research offer some insights (Seidel et al., 2021), there is a limit to what we can learn from this. Experimental studies provide causal leverage but often struggle with external validity, assuming that actors are passively exposed to evidence. In practice, exposure to information is rarely exogenous—it is likely filtered through institutional, political, and cognitive pathways (Majone, 1989).

At the same time, the digitalization of policymaking—and policymakers' lives—has resulted in the generation of alternative data sources. Increasingly, signals about how policymakers encounter and engage with evidence providers and information are being recorded—and in principle, these traces are computationally traceable as *meaningful measures* (Lazer et al., 2021). These developments open the door to new empirical strategies for studying the evidence-policy nexus—moving beyond what actors say they do, to what their behavior and materialized choices reveal.

This dissertation takes advantage of the opportunities created by these new data environments. Building on this shift, the project turns attention away from downstream questions—such as whether policymakers update their beliefs when presented with evidence—and toward upstream matters: how, when, and under what conditions they encounter that evidence in the first place. In doing so, it contributes to a growing rethinking of the evidence–policy interface, not as a neat pipeline from research to decision, but as a complex ecosystem shaped by selective exposure, institutional filters, and the cognitive pressures of policy work. This perspective not only invites new questions but also enables new methods—leveraging unstructured data to trace how evidence enters, circulates, or is filtered in and out of policy processes. In the next section, I outline the four papers that make up this dissertation, each exploring a different facet of that broader puzzle.

New data, old questions: Empirical windows into elite–evidence interactions

Despite the logical sequence whereby exposure to evidence is a precondition for its use, the empirical literature has tended to focus more on how policymakers respond to evidence than on how they encounter it. This imbalance partly reflects the methodological limitations researchers face when trying to observe policymaker behavior in real-world contexts. Much of what we know about the evidence–policy interface comes from in-depth qualitative studies using participant observation or elite interviews within single institutions or cases (e.g., Bogenschneider, Day, and Bogenschneider, 2021; Geddes, 2021; Geddes, 2024). Another common strategy is to rely on self-reported data from surveys (e.g., Weiss and Bucuvalas, 1980; Seidel et al., 2021), which often suffer from low response rates, likely social desirability bias, and limited generalizability—especially in elite settings, where access and trust are hard to secure. Getting a legislator to fill out a survey is a valuable accomplishment in itself; doing so at scale, or repeatedly over time, is even more difficult.

These methods provide in-depth and context-rich evidence points on how policymakers approach evidence and interact with research communities. But they also come with trade-offs: limited scalability, uncertain representativeness, and a reliance on retrospective accounts. As a result, we still know little about patterns that relate to how decision-makers encounter expert knowledge across broader systems or in more routine settings. In particular, existing studies often overlook what happens before evidence is formally consulted, cited, or used—what could be called the exposure 'in the wild' to experts and expertise that takes place in the background of policy work.

Recent developments in digital communication and data infrastructure offer new opportunities to address this gap. Increasingly, the everyday activities of policymakers—who they follow, respond to, or engage with online—leave behind behavioral traces that are observable and computationally traceable. While these traces do not capture internal thought processes or intentions, they do provide a unique window into patterns of attention and potential exposure. Using such data enables researchers to study elite–evidence interactions on a larger scale, over time, and across contexts, thereby opening up complementary empirical strategies that can enrich, rather than replace, qualitative and experimental insights. The first paper in this dissertation takes up this opportunity.

Leveraging behavioral traces from the digital space. The first paper, *Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience*, contributes to the broader dissertation by exploring the potential offered by digital behavioral data to study the interactions between political and scientific elites. It complements and extends traditional approaches to analyzing the evidence-policy nexus by employing large-scale, passively collected data from social media platforms. In doing so, I propose an alternative window into lawmakers' information environments—one that allows us to detect, at scale, their interactions with academic researchers and infer patterns of engagement that would otherwise remain opaque.

Using an original dataset of over 3,600 legislators across 12 democracies, this study assesses the degree to which lawmakers engage with researchers in their online networks by linking their digital footprints on Twitter (now X) to a large global database of academic researchers. This work contributes to a growing body of literature that views social media as a crucial site for political and scientific communication (Jungherr, 2016; Mede et al., 2024). It builds on findings that social media platforms—particularly Twitter—have not only been central tools for political actors to interact with constituents and share views (Castanho Silva and Proksch, 2022), but also for researchers to disseminate their work and build influence outside academia (Klar et al., 2020; Garg and Fetzer, 2025). As such, this paper takes seriously the proposition that online interactions can reveal individuals' 'offline' dispositions (Barberá, 2015; He and Tsvetkova, 2023).

Crucially, this digital data source enables scalable, cross-national analysis of legislators' engagement with research—a task that is often prohibitively resource-intensive using interviews, surveys, or active participation. This comparative reach, and scalability, helps address a notable gap in the literature, which has historically been dominated by studies of lawmakers in the U.S. and other Anglophone contexts (Ouimet et al., 2023). This study expands our understanding of how institutional, cultural, and partisan factors shape the interface between science and politics by examining behavioral patterns across a diverse set of democracies.

In the study, I find that these two elite groups—legislators and scientists—do in fact encounter each other in the digital realm. Still, visible engagement is relatively rare. Legislators follow, mention, retweet, and reply to researchers at low rates, but they do. Additionally, their online behaviors are concentrated among a narrow set of scholars, primarily social scientists. These patterns from the digital space offer a peek into the everyday information practices of elected officials.

The paper contributes to the dissertation's overarching goal of unpacking how exposure to evidence can occur by highlighting the structural and individual-level factors that relate to digital encounters between political and scientific elites. Legislators with research experience (e.g., those holding PhDs) and members of Green parties are more inclined to follow and interact with researchers. In contrast, members of conservative and radical right parties are significantly less likely to engage with academic content (and academics). These findings suggest that the willingness to seek out or amplify expert voices is not uniformly distributed across the political spectrum or legislative career paths—it is conditioned by ideological and professional predispositions that may mirror deeper differences in epistemic trust and openness to scientific input.

Importantly, this work also explores the potential for shifts in legislator behaviors in response to changes in issue salience and information needs in the context of the outset of the COVID-19 pandemic. The results reveal a marked increase in engagement with researchers, particularly in fields tied to public health and medical science, during the early stages of the pandemic. This responsiveness suggests that digital behaviors are not static but can shift in moments of crisis when evidence becomes more politically and socially salient.

This paper contributes to the broader dissertation by illustrating how novel data sources can shed light on one of the most elusive stages in the knowledge-to-policy process: the moment of potential exposure. By tracking when, how, and with whom legislators engage online, I move beyond stated preferences or hypothetical scenarios and capture behaviors as they occur in a granular manner. While these digital traces cannot capture the full complexity of policymaking processes, they offer scalable, cross-contextual indicators that complement more traditional approaches and help map out the information ecosystem in which policymakers operate.

Taken together, this paper highlights the potential of using digital behavioral data to study elite interactions with research and expertise. In combination with the other studies in this dissertation, it lays the groundwork for a more comprehensive understanding of how knowledge can make its way into the policy process. It emphasizes that before research can be used, it must first be seen, recognized, and deemed credible.

Tracing global patterns of evidence use in policy documents. The second paper, *Policy documents across 185 countries predominantly rely on evidence from the Global North*, co-authored with Roman Senninger, contributes to the broader dissertation by systematically examining the geography of *scholarly* and *policy-based* evidence referenced in policymaking worldwide. It extends prior work examining referencing practices that often focus narrowly on scientific citations to academic work in specific subdisciplines, single policy areas, or predominantly U.S.-centric contexts (Haunschild and Bornmann, 2017; Isett, Hicks, and Kingsley, 2024; Furnas, LaPira, and Wang, 2025). We build and analyze one of the most comprehensive datasets to date, covering multiple policy domains and nearly all sovereign states. Our analyses draws on over 1.2 million government policy documents from 185 countries, capturing approximately 3.5 million citations to scholarly works and 740,000 citations to policy sources—including government agencies, academic researchers, international organizations, and think tanks—offering a peek into the *documented*, *accessible*, and *digitally visible* evidence that informs policymaking globally.

Building on the dissertation's overarching theme, this paper uncovers signals of evidence flows into the policy world by employing novel data, opening new empirical windows into elite–evidence interactions. Specifically, we provide an alternative vantage point into the complex information environments shaping policymaking by examining where governments source their knowledge and how these knowledge flows are geographically distributed and structured through one of the most established institutional vehicles of knowledge transfer in the policy domain—the policy document.

To conduct this analysis, we constructed a relational database integrating data from Overton (Szomszor and Adie, 2022) and OpenAlex (Priem, Piwowar, and Orr, 2022), two of the largest and most comprehensive bibliometric and policy citation data platforms currently available. Using millions of citations extracted from the policy documents, we map cross-national citation patterns to identify: (a) the relative prominence of domestic versus foreign evidence within policy texts, (b) which countries' policy and scientific outputs achieve the greatest reach, and (c) how citation patterns vary across distinct policy domains with differing knowledge needs and demands. This empirical investigation speaks directly to a growing interdisciplinary literature on global knowledge hierarchies (Nielsen and Andersen, 2021; Castro Torres and Alburez-Gutierrez, 2022; Gomez, Herman, and Parigi, 2022), addressing a critical yet underexplored question about how spatial and institutional geographies of knowledge production and circulation manifest concretely in policymaking arenas worldwide.

Our results reveal several important findings. Governments in the Global South tend to rely disproportionately on foreign materials, especially evidence originating from the global North. At the same time, higher-income and more research-intensive countries predominantly draw on domestic expertise and institutions. Among the most frequently cited producers of both policy and academic research are countries in the global North, including the U.S., U.K., and member states of the European Union, underscoring their continued dominance in the global knowledge sphere and their relative importance in shaping international policy conversations. While some policy domains exhibit relative variation in the balance between scholarly and policy-based evidence, suggesting differential epistemic cultures and knowledge needs, the overarching pattern of global North predominance remains consistent across domains. These findings highlight persistent asymmetries in access to and influence over research knowledge, which carry important implications for the relevance, legitimacy, and contextual appropriateness of evidence informing policymaking in diverse national settings.

By focusing on such large-scale, cross-national patterns, this paper advances the dissertation's overarching goal of exploring likely settings and vehicles of exposure to expert knowledge, revealing which forms of knowledge are made visible and are likely to come to the attention of policy decision-makers around the world. Our findings highlight the critical role of knowledge infrastructures, institutional visibility, and geopolitical power relations in mediating the reach of evidence, emphasizing that patterns of referencing are embedded within broader social, economic, and political contexts that shape knowledge flows.

Taken together, the results from this paper contribute to a more nuanced understanding of the knowledge-to-policy interface by documenting the global distribution of evidence cited in official policy documents. It presents a detailed characterization of the knowledge ecosystem policymakers navigate as part of their work. This research is particularly relevant given the current state of affairs in some of these global reference points, where academic and bureaucratic systems that produce a substantial portion of the knowledge used worldwide face increasing challenges and contestation. Our findings thus carry important implications for ongoing discussions about evidence generation and dissemination, global equity, and the resilience of knowledge systems under pressure.

Building up infrastructure to study expertise in legislative committees. The third paper, *The Bundestag Expert Witness Tracker (BEWIT): A database of German Bundestag public expert hearings*, introduces a novel empirical resource that provides a structured interface for studying expert information provision in the German Parliament. While information and expert knowledge are widely recognized as vital inputs shaping democratic decision-making (Austen-Smith, 1993; Jones and Baumgartner, 2005; Baumgartner and Jones, 2015), the specific settings in which lawmakers interact with and select expert sources remain understudied, especially compared to executive-branch processes (Oliver et al., 2014). This study aims to contribute to the field by facilitating the examination of public expert hearings in the German Bundestag—formal, recorded events where committees solicit testimony on complex or contested legislation.

Once again, speaking to the dissertation's broader theme of extracting signals from settings where elite decision-makers might encounter evidence, this paper presents the BEWIT, a newly constructed database capturing over 1,800 public committee hearings from 2009 to 2024 and documenting more than 10,000 unique expert participants. These hearings provide a rare, observable arena to study how legislators act as active curators of expertise, making strategic decisions about which voices to include in public policymaking debates (e.g., Geddes, 2018; Ban, Park, and You, 2024; Geddes, 2024). I focus on publicly accessible records, including recent procedural transparency reforms revealing which party group invited each expert, to illustrate how the BEWIT offers unprecedented granularity in studying actors in legislative knowledge environments.

Using this comprehensive database, the paper addresses key questions about expertise in legislatures: Who gets invited to testify? What kinds of scholarly knowledge and expertise are prioritized across different policy domains? Is the composition of the expert pools influenced by procedural changes, such as increased transparency of invitation sources? Two empirical applications showcase BEWIT's analytical potential. First, the paper characterizes the expert witness pool, highlighting the substantial representation of academic researchers alongside interest groups, government officials, and other stakeholders, and examines variation by policy area and committee context. Second, it zooms in on a procedural reform in the 20th legislative session that made invitation sources public (i.e., disclosing the party responsible for each expert), showing that this increased transparency correlates with a modest decline in interest group representation, suggesting that legislators might have adjusted expert selection in response to the protocol changes.

These findings reveal the multifaceted role of expertise in legislative politics, where expert testimony functions not only as informational input but also as a symbolic and strategic resource shaped by partisan dynamics and institutional norms. Beyond its substantive contributions, the BEWIT can serve as a valuable tool for future research, enabling investigations into expert diversity, academic prestige effects, partisan alignment, and the interplay of expertise and political power across legislatures.

Overall, this paper enriches the dissertation by opening a new empirical window into the informational foundations of parliamentary decision-making, providing a rigorous framework for studying who gets to speak in democratic politics and how expertise is (sometimes strategically) platformed within legislatures.

Shifting from behavior to capacity: Measuring evidence consumption literacy in policy contexts. The fourth paper, *Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory*, pivots from observing elite behavior to measuring their capacity to engage with evidence. In this collaborative project, Simon Munzert and I develop and validate the Inventory for Numeracy, Statistics, and Policy-oriented Inference and REasoning (INSPIRE)—a psychometrically robust instrument designed to assess how individuals understand, evaluate, and apply scientific evidence in complex decision-making environments.

While calls to strengthen evidence-informed policymaking abound in the policy environment (e.g., Juncker, 2014; U.S. Congress, 2019), empirical tools for diagnosing the skills required to engage with evidence have lagged. The INSPIRE instrument fills this gap by targeting applied reasoning competencies relevant to public policy, including statistical, scientific, and data literacy, as well as causal reasoning. The scale is constructed using principles from item response theory (IRT) and validated across general population and policy-adjacent professional samples.

In this paper, we offer both a diagnostic tool and a conceptual framework for assessing evidence consumption competencies in policy contexts. Unlike factual knowledge tests or generic numeracy scales, the INSPIRE inventory is tailored to the specific demands of engaging with scientific findings in practical policy settings. The paper shows that INSPIRE scores capture meaningful variation in reasoning ability that is not reducible to sociodemographic characteristics such as education or political orientation, and links these competencies to behavioral outcomes, such as selecting appropriate evidence for policy questions or rejecting scientifically inaccurate claims.

The inventory is also designed for practical deployment. The scale supports modular administration (e.g., in tailored short-form versions) and can be used as a monitoring tool, or to evaluate the effectiveness of interventions, making it a useful tool for both researchers and institutions seeking to build evidence capacity. In this sense, the INSPIRE inventory complements the institutional lens of earlier papers by shifting attention to the individual-level skills that underpin effective evidence use.

Outlook

The findings across this dissertation reveal that interactions between elite policy decision-makers and evidence are far from straightforward—they are nuanced, shaped by political, institutional, and individual-level factors. Notably, characteristics of policymakers, particularly their political affiliations and ideological orientations, significantly influence their engagement with academic research providers. This dynamic is evident in both social media traces and decisions about which experts are platformed in legislative settings.

Across the evidence-policy interface, social scientists emerge as the most prominently featured group of academic experts. Their work consistently appears more frequently, whether in social media engagement behaviors, the evidence cited in policy documents across the world, or the expert witnesses invited to testify in the Bundestag. This pattern reflects broader global dynamics, where knowledge emanating from established epistemic reference points dominates policymaking discourse worldwide.

At the individual level, the capacity to engage with evidence matters: those with higher evidence consumption literacy are better equipped to select relevant and suitable evidence amid the cognitive and attentional constraints inherent to the busy policy environment. This highlights the importance of cognitive skills in shaping evidence use beyond institutional factors.

Many questions remain open, and the extensive data infrastructure developed here offers a foundation for exploring new facets of the knowledge-to-policy interface. For example, analyzing the text content of social media communications can help investigate long-standing hypotheses about the relationship between scientists and policymakers, such as Weiss' typology of evidence use models (Weiss, 1979) and Caplan's "two communities" thesis (Caplan, 1979). Future research could explore whether communication styles and language in digital spaces reflect these typology or indicate distinct discursive patterns between groups.

While this project took a macro and cross-national perspective—particularly with policy document data—zooming in on specific actors and contexts remains a promising avenue. Collaborative work examining international organizations as evidence providers for the European Union (Minetto, Ramirez-Ruiz, and Senninger, 2025) exemplifies such an approach, and many other questions await investigation.

The Bundestag Expert Witness Tracker (BEWIT) is designed for ongoing growth, with plans to incorporate richer data formats—such as the parsed verbatim protocols—to extend analyses of how legislators engage with experts with higher granularity. Similarly, the INSPIRE inventory holds potential beyond research as a monitoring and assessment tool for interventions aimed at strengthening evidence engagement among policymakers and policy professionals. Building partnerships with practitioners interested in applying these insights could help translate academic knowledge into practical improvements in evidence-informed policymaking.

Together, these future directions can advance our collective understanding of the complex, multi-layered relationship between evidence and elite decision-making. In doing so, they can help unpack the evidence ecosystem that shapes policy and, ultimately, support the development of a better-equipped policy workforce capable of tackling increasingly complex societal challenges.

References

Simon, Herbert A. (1944). "Decision-Making and Administrative Organization". *Public Administration Review* 4.1, p. 16.

Lindblom, Charles E. (1959). "The Science of "Muddling Through"". *Public Administration Review* 19.2, pp. 79–88.

Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". *Science* 185.4157, pp. 1124–1131.

Janis, Irving L. and Leon Mann (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Decision making: A psychological analysis of conflict, choice, and commitment. New York, NY, US: Free Press.

Weiss, Carol H. (1977). "Research for Policy's Sake: The Enlightenment Function of Social Research". *Policy Analysis* 3.4, pp. 531–545.

Caplan, Nathan (1979). "The Two-Communities Theory and Knowledge Utilization". *American Behavioral Scientist* 22.3, pp. 459–470.

Weiss, Carol H. (1979). "The Many Meanings of Research Utilization". *Public Administration Review* 39.5, pp. 426–431.

Weiss, Carol H. and Michael J. Bucuvalas (1980). "Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research". *American Sociological Review* 45.2, pp. 302–313.

Majone, Giandomenico (1989). *Evidence, argument, and persuasion in the policy process*. Nachdr. Politics. New Haven: Yale Univ. Press.

Austen-Smith, David (1993). "Information and Influence: Lobbying for Agendas and Votes". *American Journal of Political Science* 37.3, pp. 799–833.

Ham, Chris, David J. Hunter, and Ray Robinson (1995). "Evidence based policymaking". *BMJ* 310.6972, pp. 71–72.

Davies, Huw T. O., Sandra M. Nutley, and Peter C. Smith (1999). "Viewpoint: What Works? The Role of Evidence in Public Sector Policy and Practice". *Public Money & Management* 19.1, pp. 3–5.

Sanderson, Ian (2002). "Evaluation, Policy Learning and Evidence-Based Policy Making". *Public Administration* 80.1, pp. 1–22.

Jones, Bryan D. and Frank R. Baumgartner (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.

Head, Brian W. (2008). "Three Lenses of Evidence-Based Policy". *Australian Journal of Public Administration* 67.1, pp. 1–11.

Contandriopoulos, Damien, Marc Lemire, Jean-Louis Denis, and Émile Tremblay (2010). "Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature". *The Milbank Quarterly* 88.4, pp. 444–483.

Gilardi, Fabrizio (2010). "Who Learns from What in Policy Diffusion Processes?" *American Journal of Political Science* 54.3, pp. 650–666.

Nevo, Isaac and Vered Slonim-Nevo (2011). "The Myth of Evidence-Based Practice: Towards Evidence-Informed Practice". *The British Journal of Social Work* 41.6, pp. 1176–1197.

Cairney, Paul (2012). *Understanding Public Policy: Theories and Issues*. Bloomsbury Academic.

National Research Council (2012). *Using Science as Evidence in Public Policy*. Washington, D.C.: National Academies Press.

Juncker, Jean-Claude (2014). *Mission letter to Commissioner Moedas*.

Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas (2014). "A systematic review of barriers to and facilitators of the use of evidence by policymakers". *BMC Health Services Research* 14.1, p. 2.

Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". *Political Analysis* 23.1, pp. 76–91.

Baumgartner, Frank R. and Bryan D. Jones (2015). *The Politics of Information: Problem Definition and the Course of Public Policy in America*. Chicago, IL: University of Chicago Press.

Kraft, Patrick W., Milton Lodge, and Charles S. Taber (2015). "Why People "Don't Trust the Evidence": Motivated Reasoning and Scientific Beliefs". *The ANNALS of the American Academy of Political and Social Science* 658.1, pp. 121–133.

Cairney, Paul (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan UK.

Head, Brian W. (2016). "Toward More "Evidence-Informed" Policy Making?" *Public Administration Review* 76.3, pp. 472–484.

Jungherr, Andreas (2016). “Twitter use in election campaigns: A systematic literature review”. *Journal of Information Technology & Politics* 13.1, pp. 72–91.

Caramani, Daniele (2017). “Will vs. Reason: The Populist and Technocratic Forms of Political Representation and Their Critique to Party Government”. *American Political Science Review* 111.1, pp. 54–67.

Gollust, Sarah E., Jane W. Seymour, Maximilian J. Pany, Adeline Goss, Zachary F. Meisel, and David Grande (2017). “Mutual Distrust: Perspectives From Researchers and Policy Makers on the Research to Policy Gap in 2013 and Recommendations for the Future”. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 54, p. 0046958017705465.

Haunschild, Robin and Lutz Bornmann (2017). “How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data”. *Scientometrics* 110.3, pp. 1209–1216.

Kenny, Caroline, David Christian Rose, Abbi Hobbs, Chris Tyler, and Jason Blackstock (2017). *The role of research in the UK Parliament*. Vol. 1. Parliamentary Office of Science and Technology.

Parkhurst, Justin (2017). *The Politics of Evidence: From evidence-based policy to the good governance of evidence*. Taylor & Francis.

Walgrave, Stefaan and Yves Dejaeghere (2017). “Surviving Information Overload: How Elite Politicians Select Information”. *Governance* 30.2, pp. 229–244.

Baron, Jon (2018). “A Brief History of Evidence-Based Policy”. *The ANNALS of the American Academy of Political and Social Science* 678.1, pp. 40–50.

Bisgaard, Martin and Rune Slothuus (2018). “Partisan Elites as Culprits? How Party Cues Shape Partisan Perceptual Gaps”. *American Journal of Political Science* 62.2, pp. 456–469.

Geddes, Marc (2018). “Committee Hearings of the UK Parliament: Who gives Evidence and does this Matter?” *Parliamentary Affairs* 71.2, pp. 283–304.

Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen (2019). “The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians”. *British Journal of Political Science* 49.3, pp. 1117–1140.

Banuri, Sheheryar, Stefan Dercon, and Varun Gauri (2019). “Biased Policy Professionals”. *World Bank Economic Review* 33.2, pp. 310–327.

Bowers, Jake and Paul F. Testa (2019). “Better Government, Better Science: The Promise of and Challenges Facing the Evidence-Informed Policy Movement”. *Annual Review of Political Science* 22.1, pp. 521–542.

Mair, David, Laura Smillie, PLACA Giovanni La, Florian Schwendinger, Milena Raykovska, Zsuzsanna Pasztor, and BAVEL Rene Van (2019). *Understanding our Political Nature: How to put knowledge and reason at the heart of political decision-making*.

U.S. Congress (2019). *Foundations for Evidence-Based Policymaking Act of 2018*. H.R. 4174, 115th Congress (2017–2018). U.S. Congress.

Klar, Samara, Yanna Krupnikov, John Barry Ryan, Kathleen Searles, and Yotam Shmargad (2020). “Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work”. *PLOS ONE* 15.4, e0229446.

Tollefson, Jeff (2020). “How Trump damaged science — and why it could take decades to recover”. *Nature* 586.7828, pp. 190–194.

Algan, Yann, Daniel Cohen, Eva Davoine, Martial Foucault, and Stefanie Stantcheva (2021). “Trust in scientists in times of pandemic: Panel evidence from 12 countries”. *Proceedings of the National Academy of Sciences* 118.40, e2108576118.

Bogenschneider, Karen, Elizabeth Day, and Bret N. Bogenschneider (2021). “A window into youth and family policy: State policymaker views on polarization and research utilization”. *The American Psychologist* 76.7, pp. 1143–1158.

Crowley, D. Max, J. Taylor Scott, Elizabeth C. Long, Lawrie Green, Azaliah Israel, Lauren Supplee, Elizabeth Jordan, Kathryn Oliver, Shannon Guillot-Wright, Brittany Gay, Rachel Storace, Naomi Torres-Mackie, Yolanda Murphy, Sandra Donnay, Jenna Reardanz, Rebecca Smith, Kristina McGuire, Elizabeth Baker, Ana Antonopoulos, Mary McCauley, and Cagla Giray (2021). “Lawmakers’ use of scientific evidence can be improved”. *Proceedings of the National Academy of Sciences* 118.9, e2012955118.

Geddes, Marc (2021). “The webs of belief around ‘evidence’ in legislatures: The case of select committees in the UK House of Commons”. *Public Administration* 99.1, pp. 40–54.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini (2021). “How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities”. *American Economic Review* 111.5, pp. 1442–1480.
Lazer, David, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford (2021). “Meaningful measures of human society in the twenty-first century”. *Nature* 595.7866, pp. 189–196.
Mehmood, Sultan, Shaheen Naseer, and Daniel L Chen (2021). “Training Policymakers in Econometrics”.
Nakajima, Nozomi (2021). “Evidence-Based Decisions and Education Policymakers”.
Nielsen, Mathias Wullum and Jens Peter Andersen (2021). “Global citation inequality is on the rise”. *Proceedings of the National Academy of Sciences* 118.7, e2012208118.
Seidel, Katja, Hans Verbeek, Sophie Fessel, and Florian Meer (2021). *Nutzen von wissenschaftlicher Evidenz: Erwartungen an wissenschaftliche Expertise : Bericht über eine Bundestagsbefragung im Rahmen der Leopoldina-Evidenzinitiative unter Leitung von Claudia Buch, Regina T. Riphahn und Monika Schnitzer*. Tech. rep. MyCoRe Community, vol. 27.
Castanho Silva, Bruno and Sven-Oliver Proksch (2022). “Politicians unleashed? Political communication on Twitter and in parliament in Western Europe”. *Political Science Research and Methods* 10.4, pp. 776–792.
Castro Torres, Andrés F. and Diego Alburez-Gutierrez (2022). “North and South: Naming practices and the hidden dimension of global disparities in knowledge production”. *Proceedings of the National Academy of Sciences* 119.10, e2119373119.
Gomez, Charles J., Andrew C. Herman, and Paolo Parigi (2022). “Leading countries in global science increasingly receive more citations than other countries doing similar research”. *Nature Human Behaviour* 6.7, pp. 919–929.
Lee, Nathan (2022). “Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers”. *American Political Science Review* 116.2, pp. 677–688.
Priem, Jason, Heather Piwowar, and Richard Orr (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*.
Szomszor, Martin and Euan Adie (2022). “Overton: A bibliometric database of policy document citations”. *Quantitative Science Studies* 3.3, pp. 624–650.
Geddes, Marc (2023). “Good Evidence: How do select committees use evidence to support their work?”
He, Yuanmo and Milena Tsvetkova (2023). “A Method for Estimating Individual Socioeconomic Status of Twitter Users”. *Sociological Methods & Research*, p. 00491241231168665.
Ouimet, Mathieu, Morgane Beaumier, Adrien Cloutier, Alexandre Côté, Éric Montigny, François Gélineau, Steve Jacob, and Stéphane Ratté (2023). “Use of research evidence in legislatures: a systematic review”. *Evidence & Policy* -1.aop, pp. 1–18.
Vivalt, Eva and Aidan Coville (2023). “How do policymakers update their beliefs?” *Journal of Development Economics* 165, p. 103121.
Ban, Pamela, Ju Yeon Park, and Hye Young You (2024). *Hearings on the Hill: The Politics of Informing Congress*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press.
Geddes, Marc (2024). “Comparing evidence use in parliaments: the interplay of beliefs, traditions, and practices in the UK and Germany”. *Policy and Society* 43.4, pp. 447–462.
Isett, Kimberley R., Diana Hicks, and Gordon Kingsley (2024). “Blending the two communities: concurrent evidence use in state-level policy processes”. *Public Management Review* 0.0, pp. 1–18.
Mede, Niels G., Viktoria Cologna, Sebastian Berger, John Besley, Cameron Brick, Marina Joubert, Edward Maibach, Sabina Mihelj, Naomi Oreskes, and Mike S. Schäfer (2024). *Public communication about science across 68 countries: Global evidence on how people get information and communicate about science-related matters*.
Senninger, Roman and Henrik B. Seeberg (2024). “Which Information Do Politicians Pay Attention To? Evidence from a Field Experiment and Interviews”. *British Journal of Political Science*, pp. 1–18.
Cologna, Viktoria et al. (2025). “Trust in scientists and their role in society across 68 countries”. *Nature Human Behaviour*, pp. 1–18.
Furnas, Alexander C., Timothy M. LaPira, and Dashun Wang (2025). “Partisan disparities in the use of science in policy”. *Science* 388.6745, pp. 362–367.
Garg, Prashant and Thiemo Fetzer (2025). “Political expression of academics on Twitter”. *Nature Human Behaviour*, pp. 1–18.

Minetto, Francesca, Sebastian Ramirez-Ruiz, and Roman Senninger (2025). "International Organizations as Evidence Sources for the European Union".
Vivalt, Eva, Aidan Coville, and Sampada Kc (2025). "Local knowledge, formal evidence, and policy decisions". *Journal of Development Economics* 174, p. 103425.

Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience

Sebastian Ramirez-Ruiz (Hertie School)*

Abstract. Interactions between the policy and academic communities can play an important role in political decisionmaking. Still, the fact that much of the policymaking process happens behind closed doors obscures our understanding of the relationships between political decision-makers with academic researchers. To address this challenge, this paper introduces a novel approach that leverages online behavioral data from social media to examine how legislators interact with researchers. By analyzing data from 3,670 lawmakers in 12 countries merged to a novel database of 410K academic researchers on Twitter, this study provides new insights into these otherwise hidden interactions. The findings suggest that lawmakers do follow, yet rarely visibly engage with researchers online. Lawmakers from conservative and radical right parties follow and engage less with researchers online than their colleagues from other parties. While the base engagement is relatively low across legislatures, it can increase when expertise gains salience. During the early stages of the COVID-19 pandemic, marked by policy uncertainty involving a novel and technically complex policy issue, lawmakers' overall inclination to follow and engage with scholars increased, most prominently targeting researchers from the medical sciences. These findings offer new insights into when and how lawmakers publicly attend to academic expertise, contributing to a broader understanding of political elites' symbolic and informational engagement with science.

Keywords. legislative elites | academic researchers | elite digital traces | social media

*Corresponding author. Address: Friedrichstrasse 180, 10117 Berlin, Germany. Email: ramirez-ruiz@hertie-school.org, Web: seramirezruiz.github.io.

Global challenges such as climate change and the COVID-19 pandemic underscore the potential importance of academic researchers, scientific research and insights on policy formation (Bavel et al., 2020; Berger et al., 2021). Arguably, research evidence can be valuable for policy decisions by presenting information to create and execute policies to address societal problems. At the same time, we live in a period with an unprecedented expansion of the scientific workforce, accompanied by a surge in research output (Heuer, Einaudi, and Kang, 2023; Fortunato et al., 2018; Bornmann, Haunschild, and Mutz, 2021). The increasingly growing pool of academic experts and knowledge converges with the reality that policymakers receive a steady flow of information on societal issues from diverse sources (Jones and Baumgartner, 2005; Senninger and Seeberg, 2024). Consequently, the decisions policymakers make about engaging with researchers and evidence—and ultimately whom they pay attention to—can have significant implications.

There is a long-standing research tradition in the social sciences looking at the interactions between expert academic communities and public decision-makers (Weiss, 1979; Caplan, 1979; Huberman, 1994; Haas, 1992). However, there are gaps in our understanding about whether actors in the policy sphere engage with researchers and are exposed to and consume evidence (National Research Council, 2012). One of the main obstacles to studying this is the confidentiality and unobservability of elite behaviors. Much of the previous work on this domain relies on participant observation and semi-structured interviews (Geddes, 2021; Bogenschneider, Day, and Bogenschneider, 2021), self-reports (Caplan et al., 1975; Weiss and Bucuvalas, 1980), and information provision experiments (Vivalt and Coville, 2023; Lee, 2022; Baekgaard et al., 2019). While these approaches have deepened our understanding of how policymakers encounter and use expertise, they often offer limited insight into public-facing signals of engagement. In this paper, I propose a complementary approach that uses digital behavioral data from social media to observe whether legislators publicly engage with researchers. This offers a novel lens on the visibility and conditions under which political elites acknowledge or interact with academic experts in the public sphere.

In this paper, I investigate the legislator relationships to academic researchers on Twitter (now X), exploring potential variation across different political contexts, jurisdictions, and individuals. I propose using online behavioral data from social media as a means to study legislators' relationships and information flows with researchers, with the aim to overcome the unobservability of their behaviors. I argue that online signals of engagement with researchers can reflect their overarching disposition to consider academics as a legitimate and valuable source of information.

Equipped with granular social media data from legislators across 12 democracies, I tackle three overarching sequential questions: 1. Do we observe lawmakers following and engaging with researchers 'in the wild'? 2. Are there legislator and legislature-level factors that help predict online engagement with researchers? And, 3. do legislators adapt their online behaviors when expertise gains salience?

My analyses suggest that (i) researchers are part of legislators' social media networks, but legislator-researcher online encounters are rare events, (ii) some legislator characteristics, importantly their political affiliation and educational levels, are related with engagement, and (iii) in an instance of high demand for research evidence, such as a the COVID-19 pandemic, we observe an increase in legislators' inclination to follow and engage with experts holding specific domain knowledge.

Studying political and scientific elites on social media

Online social media platforms are increasingly important around the globe. These media have transformed market practices, information diffusion and consumption, and political communication (Lamberton and Stephen, 2016; Westerman, Spence, and Van Der Heide, 2014; Jungherr, 2016). These platforms also hold particular relevance for both political and scientific communication, offering actors in both spheres new ways to connect with broader publics and each other (Brainard, 2022).

A recent global study across 68 countries finds that social media are now the most significant source of science-related information in most regions (Mede, Cologna, et al., 2024). If this is where much of the public encounters scientific knowledge, it becomes increasingly timely to ask whether lawmakers also encounter, and publicly engage with, researchers in these spaces. Unlike traditional elite communications, which are often private, curated, or constrained, social media offer a comparatively open and observable venue where selective forms of engagement can be systematically studied.

For lawmakers, social media represent a low-cost channel to convey information to the public and appeal to constituents. These media provide flexibility to communicate, receive feedback, and gather information without the constraints of other political arenas, such as regulated debates, prepared statements, and news coverage (Castanho Silva and Proksch, 2022). For academics, social media represent a tool to advance their careers, expand knowledge networks, and scale research dissemination to reach a wider audience (Garg and Fet-

Figure 1 provides an overview of the study population and data structure. All-in-all, roughly 90% of the legislators in these democracies were on Twitter. For further information and methodological detail about the data collection, see *SI Appendix A*.

I employ the list of researcher accounts to explore the instances where legislators engage with researchers in the platform. With these newly compiled data, I contribute to past work on the relationship between academic researchers, evidence, and legislators in three ways.

First, I employ a set of behavioral measures to gauge legislators' engagement with academic researchers in the digital realm. Traditionally, studies in this area have relied on self-reported measures (Avey and Desch, 2014; Riphahn and Schnitzer, 2022; Seidel et al., 2021). While asking MPs how much they engage with academic researchers in their parliamentary activities is valuable (Dodson, Geary, and Brownson, 2015; Purtle et al., 2018), these measures can be costly and difficult to scale beyond one context.

I study six specific behaviors, each giving different affordances to users and shedding light on diverse facets of the Twitter interaction landscape:

- **Following:** This behavior involves users subscribing to the updates of other accounts on the platform, reflecting their interest in specific users.
- **Mentioning:** This is the behavior by which users reference an account in their original posts. It entails direct engagement and communication with that user.
- **Retweeting:** This action entails users forwarding posts made by other accounts to their own followers, potentially amplifying the reach of specific content.
- **Quote Tweeting:** This function allows users to share another account's post while adding their commentary, potentially contributing their perspective to the content.
- **Replying:** When a user responds to posts made by other users. It can reflect direct engagement in conversations and discussions taking place on the platform.
- **Liking:** By using the 'like' function, users generally express positive sentiments or approval for particular tweets, potentially contributing to the overall sentiment and visibility of content.

These behaviors provide indicators that can help understand the interactions between legislators and researchers on Twitter.

Second, leveraging the scalability of these measures, I extend the analyses to lawmakers from 12 countries. The majority of previous studies exploring research evidence in legislatures present insights from single-country studies, with a lopsided focus on the U.S. context (Ouimet et al., 2023). This international perspective enables us to better understand the connections between legislators and academic research providers across diverse political, institutional, and cultural contexts.

By including legislators from various countries, I am able to examine correlates at the individual and legislature levels, such as legislators' professional backgrounds, political party affiliations, and political system characteristics. As such, I assess patterns and factors related to engagement with researchers across country contexts.

Third, I examine the responsiveness of legislators to changes to the salience of expertise. The COVID-19 pandemic serves as an intriguing case study due to the heightened crisis, uncertainty, and urgency it brought about, coupled with a pressing need for guidance from scientific experts.

Existing research has shown that epistemic communities, or groups of experts with specialized knowledge, can exert significant influence in situations involving novel and technically complex policy issues, mainly when decision-makers have limited understanding of these issues (Haas, 1992; Dunlop, 2017). Crises such as the pandemic could, in principle, generate 'problem uncertainty', where policymakers are forced to grapple with uncertainty about the nature of the policy problems they face. These trends have been observed through other proxies policy environment (Yin et al., 2021). If such changes can also be observed in legislators' digital behaviors, this could emphasize that these interactions hold signals of lawmakers' inclination to engage with the academic community as an information source.

I employ the expected changes in salience of some types of expertise to investigate whether there were discernible shifts in the digital engagement behaviors of legislators with researchers in times of crisis when the stakes are high and the demand for evidence should be pronounced.

Results

Do legislators follow and engage with researchers online? The general descriptive statistics offer insights into the dynamics and interactions that characterize their relationships on Twitter.

The majority of legislators follow at least one researcher account from the list (about 86% of legislators). The median legislator was following 990 accounts of which 8 were academic researchers (see *SI Appendix, Table B1* for an overview of the distribution of these legislators do connect with researchers, they do so far less frequently than these other groups.

In addition to that, these behaviors are aimed at a select group of researchers. While 6.7% of researchers in the list of 410K appear in legislator networks, as behaviors become more active the group reduces between 1% and 4%. Furthermore, researchers in the social sciences account consistently for about half of the legislator-researcher matches across behaviors, suggesting that legislators tend to pay attention to them to a higher degree.

Are there any predictors of engagement with researchers based on specific contextual and legislator characteristics? Previous research has identified individual-level correlates of trust and politicization of science, as well as preferences for the role of scientific experts in policy debates (for an overview see Rutjens, Heine, et al., 2018). Overall, findings point to asymmetries related to political ideology (Gauchat, 2012; Lewandowsky and Oberauer, 2016; Linden et al., 2021; Funk et al., 2019), religiosity and spirituality (Rutjens, Sutton, and Lee, 2018; Rutjens, Sengupta, et al., 2022), knowledge about science and education level (Rutjens, Sutton, and Lee, 2018; Mede and Schäfer, 2020), age (Anderson et al., 2012), and gender (Gauchat, 2012; Roten, 2004). More recently, cross-national evidence of science skepticism in 24 countries suggests that there is variation of predictors across politicized scientific domains, in addition to heterogeneity in the levels of skepticism across countries (Rutjens, Sengupta, et al., 2022).

When mapping these features onto the available legislator behavioral data from Twitter, I find some differences between legislator characteristics and across legislatures. Figure 3 presents the results from a linear mixed-effect model exploring predictors of the proportion of researchers in legislators' networks. The general direction of the predictors is parallel in all the behaviors of the study (see *SI Appendix, Figures C1-C5* for the models of the remaining engagement behaviors and *Table B10* for model output).

While there are some national-level differences as evidenced by the general overview in Figure 1b, there are no discernible regional differences between legislators. Although legislators exhibit a similar propensity to follow and engage with scientists in the list across the different behaviors, the disparity becomes evident in the magnitude of their engagement. Latin American legislators, on average, follow and engage with a smaller absolute number of academic researchers compared to their counterparts from other regions. A potential explanation for the regional patterns can be on the researcher supply side and their

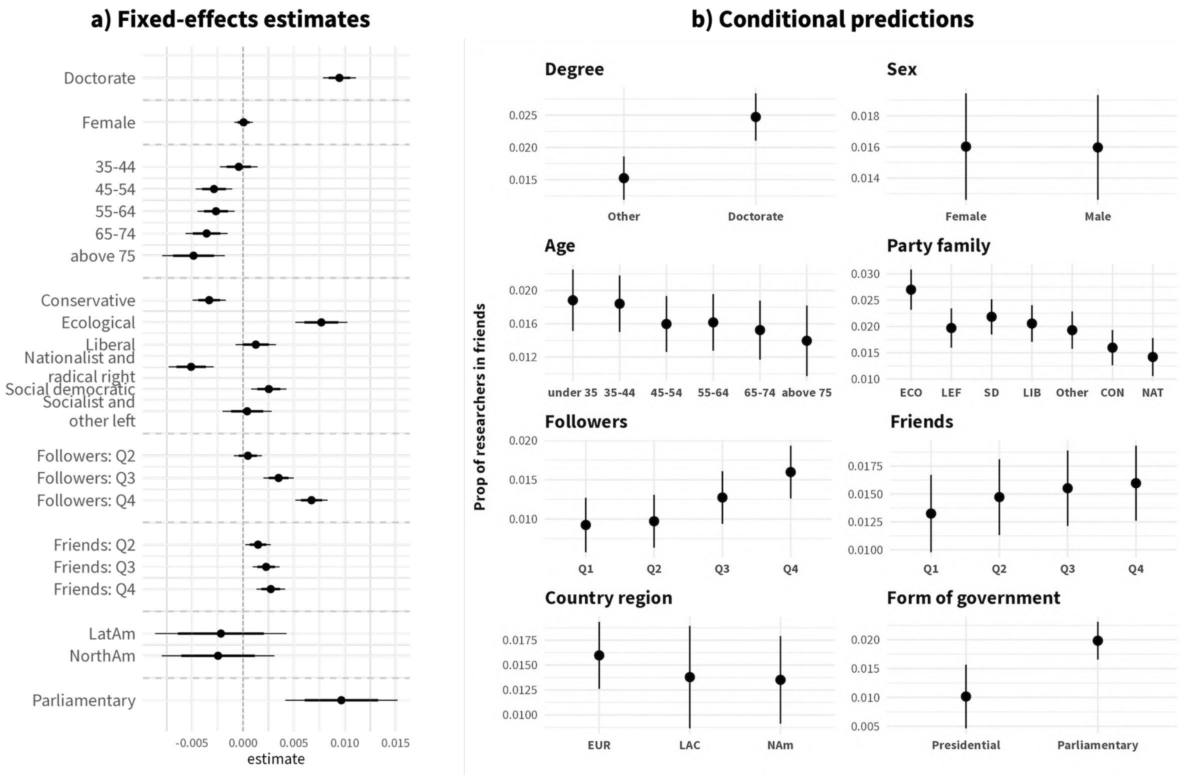


Figure 3: Estimated effects of legislator and legislature characteristics on the proportion of researchers in their networks. Results from a linear mixed-effects model with country random effects with age (under 35), party family (other), country region (Europe), system (presidential), and Q1 for followers and friends as references for categorical variables. Number of observations: 3,381; Groups: 12. Panel a presents the coefficients with 80% and 95% confidence intervals. The conditional predictions are computed with numeric covariates are held at their means and the other covariates at their modes: no research degree, presidential, European, male, 45-54, Q1, and Conservative party. The std. followers and friends are referenced to individual's legislature mean.

proximity to the legislators (see *SI Appendix, Table B3* for an overview of the geographical distribution of the researchers).

When it comes to one of the most salient and studied correlates of attitudes towards science in the general public, political ideology, these data reveal a picture consistent with previous findings (Gauchat, 2012; Lewandowsky and Oberauer, 2016; Mede and Schäfer, 2020). Legislators representing parties with right-wing and populists ideologies showcase less interest in researchers online. While legislators from most party families exhibit similar behaviors on average, legislators belonging to conservative, as well as nationalist and radical right parties are less likely to follow and engage with academic researchers and they do so at a lower rate. The opposite is true for legislators belonging to green parties. These legislators are more likely to follow and engage with researchers and do so proportionally more. Around 99% of 'greens' follow researchers in contrast to 66% for nationalist and radical right legislators.

Among the legislators in the sample on Twitter, 9% hold, or were enrolled in programs, leading to an advanced research qualification, such as doctorates or equivalent. Legisla-

P=<0.001); more likely to be targeted at a scholar compared to new edges created in the pre-COVID period.

These findings present evidence that legislators' inclination to follow and engage with researchers increased in the period when expertise from epistemic communities gained salience. Notably, according to subgroup analyses, the increases are not observed for populist left and right legislators (see *SI Appendix, Figure C13 and C14* for the party family and education subgroup marginal effects).

Although previous evidence suggests that COVID research transcended disciplinary boundaries, there are signs that at the initial stages particular attention was paid to biomedical research (Yin et al., 2021). Considering legislators' online behaviors at the dawn of the pandemic, we could anticipate a similar tendency.

That is, if the changes in behaviors reflected the increased interest, need for expertise, and interactions with scholars, we would expect them to be particularly targeted at researchers whose expertise is more closely related to the immediate questions emanating from the rise of COVID.

The models in Figure 5b look at the different behaviors targeting scholars in the specific groups based on their scientific branch. The subgroup that sticks out the most are the researchers belonging to the medical sciences. While the increase in the likelihood of a new following event being targeted at scholars is observable across all fields, the other behaviors present a picture where researchers from the medical field experience the largest increases in the COVID period, and in some cases, the only ones experiencing increases compared to the baseline period. For example, these results suggest that retweets observed in the COVID period were 2.3 times more likely to have a medical scientist as a source creator compared to retweets in the pre-COVID period. Although expertise and commentary from researchers in other branches could have been of great importance during this period, arguably, researchers in this subgroup are the more plausible members of the epistemic communities holding specialized knowledge for this particular situation.

Altogether, the temporal analyses of legislator behaviors on Twitter reveal an increase in engagement with researchers. The observed surge in legislator-to-researcher interactions evident across behaviors underscores the importance of specialized knowledge in informing the public discussion at the dawn of the COVID pandemic. Notably, the growth in attention to scholars in the biomedical research camp reflects the relative importance of expertise relevant to the immediate challenges posed by the pandemic.

Discussion

Idealized models often depict policymaking as a linear process of rational problem-solving, where decision-makers objectively integrate available evidence. In practice, lawmakers navigate a more complex reality. Their behavior is shaped by institutional, ideological, and electoral constraints that influence which information they seek, trust, and use (Cairney, 2016; Parkhurst, 2017; Hassel and Wegrich, 2022).

The growing paradigm of evidence-informed policymaking acknowledges these challenges while emphasizing the value of connecting scientific research with decision-making. However, research evidence is not always timely, salient, or packaged in ways that resonate with policymakers' agendas (Senninger and Seeberg, 2024; Walgrave and Dejaeghere, 2017). As researchers compete with other information sources for policymakers' attention, understanding how—and whether—lawmakers encounter academic expertise is increasingly relevant.

This study explores one dimension of that relationship: legislators' observable engagement with researchers on social media. While traditionally, these relationships may have been studied through more formal channels, such as committee meetings or research briefs, this study takes advantage of the new, more informal ways lawmakers and researchers can now connect. The results of this study offer evidence about the nature and scale of the relationship between legislators and academic researchers online, with a specific focus on the lawmakers' side of the engagement. This is particularly important because of recent findings suggesting that social media can be a significant source of scientific information across countries.

Due to the challenges to the study of these relationships as they develop in parliamentary halls, universities, research institutes, and conference rooms across the globe, I employ social media trace data from legislators' Twitter profiles and map it onto a novel dataset of close to 410K researcher producers.

I structured my inquiry following three sequential questions regarding the prevalence, correlates, and malleability of legislator engagement with academic researchers online. I bring forward the following messages.

First, legislators do engage with researchers online, but their engagement remains limited. This characterization is crucial to highlight because it underscores a significant gap between the potential for engagement and the actual engagement that is taking place. While there is no clear "analogue" gold standard for measuring this behavior, comparisons with other social media users—such as science and political journalists—indicate that journalists follow researchers at a much higher rate than legislators do.

In the pooled sample, 86% of legislators followed and between 60 and 80% engaged in one of the platform-specific behaviors towards researchers at least once. That said, the proportion of these behaviors targeted at researchers in reference to the totals is roughly 1% across the board. That is to say, that on average 1% of legislators' networks is comprised of, and engagement with content created by, researchers.

What these findings offer is evidence that engagement with researcher content is rare. Although lawmakers are very active content producers and sharers, the median legislator has only retweeted 4 tweets by researchers and quoted, mentioned, and replied once in their Twitter history.

Furthermore, these Twitter events feature a very small set of researchers from the list. For example, around 6.7% of researchers in the list are followed, 3.8% are retweeted, and 1.4% quote tweeted. These figures highlight the narrow scope of engagement and point to a potential feature of the expert information preferred by lawmakers, with social scientists being the dominant contributors to the conversations that lawmakers engage with online.

Second, the descriptive models highlight some contextual and legislator-level features that correlate with lawmakers' inclination to follow and engage with researchers online and their proportion. This offers us a clearer understanding of who is engaging and helps us identify the more likely channels for the uptake of research findings.

One key legislator-level predictor is their academic background. The legislators who were enrolled in, or hold, a doctorate degree were more likely to follow and engage with researchers from the list, in addition to having a larger proportion of scholars as targets of their Twitter behaviors. This suggests that legislators with research experience may be more inclined to engage. A conclusion from this is that legislators with research backgrounds may be more likely to be exposed to the insights and perspectives offered by academic researchers.

Additionally, the results present an ideological divide. On the one hand, lawmakers belonging to green parties are more likely to follow and engage with researchers. Also, the 'greens' who follow and engage with researchers tend to do it proportionally more than their fellow legislators. On the other hand, politicians from parties in the radical right, an ideological camp associated to science-skepticism and populism (Mede and Schäfer, 2020), exhibit the opposite inclinations. This partisan disparity in the willingness to engage with the academic community can have implications for lawmaking, with (dis)trust in science and its members potentially impacting the ability to address societal challenges effectively (Furnas, LaPira, and Wang, 2025).

Third, I leverage the variation to salience of expertise created by the COVID-19 pandemic to assess observed changes in legislators' online behaviors. The onset of the COVID-19 pandemic prompted an increase in online activity by legislators. Most engagement behaviors experienced growth during this period. Alongside the surge in activity, there was a notable increase in the likelihood that the legislator Twitter events were aimed at researchers compared to the pre-COVID baseline across the array of behaviors in the study.

The analyses of the different scientific branches revealed that scholars in the medical sciences experienced the larger increases in likelihood of following and engagement during the pandemic. These observed increases in engagement hold particular significance as they suggests that this was not merely a byproduct of overall online activity, but rather a manifestation of the particular importance that some experts whose knowledge could inform the crisis response gained. This can be taken as a signal that to some extent legislators were responsive to the information needs related to health and public safety, providing valuable insights into the salience of expertise during times of crisis. Beyond the specific pandemic case, this shift highlights that legislator–researcher engagement on social media is responsive to issue salience and societal events. It also supports the idea that social media data can reveal meaningful patterns in elite attention to expertise, even if indirect.

The study of online behavioral data provides a window into the complex relationship between academic researchers and legislators in the digital age. While social media platforms offer a potential bridge for connecting these two communities, the limited nature of legislators' online engagement with researchers suggests that more efforts are needed to facilitate meaningful interactions and knowledge sharing.

Naturally, these results cannot capture nor offer a complete overview of the complexities embedded in the interactions between lawmakers and researchers, even less so of research insights influencing a policymaker's thinking. However, the signals uncovered from the digital realm in this study do talk to a crucial condition in the potential chain of events that could lead to 'policy impact' of research findings; that is, legislators engaging with researchers and the insights they generate in the first place.

These findings shed light on some of the features related to legislators' inclination to engage with researchers, in addition to potential circumstances under which contact across the two communities can increase.

Acknowledgments

I would like to express my gratitude to Simon Munzert, Francesca Minetto, Lisa Oswald, as well as audiences at the CIVICA Political Behavior and Institutions Conference, Berlin-Brandenburg Political Behavior Workshop, Aarhus University, EPSA, and the University of Zurich, for their valuable feedback. The author acknowledges financial support by the Hertie School in the form of a doctoral stipend. The author received no specific funding for this work.

Competing interests

The author declares no competing interests.

Ethical standards

This study received ethics approval by the Hertie School Research Ethics Office (Application ID—20241111-133)

References

Anderson, Ashley A., Dietram A. Scheufele, Dominique Brossard, and Elizabeth A. Corley (2012). "The Role of Media and Deference to Scientific Authority in Cultivating Trust in Sources of Information about Emerging Technologies". *International Journal of Public Opinion Research* 24.2, pp. 225–237.

Avey, Paul C. and Michael C. Desch (2014). "What Do Policymakers Want From Us? Results of a Survey of Current and Former Senior National Security Decision Makers". *International Studies Quarterly* 58.2, pp. 227–246.

Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen (2019). "The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians". *British Journal of Political Science* 49.3, pp. 1117–1140.

Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". *Political Analysis* 23.1, pp. 76–91.

Bavel, Jay J. Van, Katherine Baicker, Paulo S. Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J. Crockett, Alia J. Crum, Karen M. Douglas, James N. Druckman, John Drury, Oeindrila Dube, Naomi Ellemers, Eli J. Finkel, James H. Fowler, Michele Gelfand, Shihui Han, S. Alexander Haslam, Jolanda Jetten, Shinobu Kitayama, Dean Mobbs, Lucy E. Napper, Dominic J. Packer, Gordon Pennycook, Ellen Peters, Richard E. Petty, David G. Rand, Stephen D. Reicher, Simone Schnall, Azim Shariff, Linda J. Skitka, Sandra Susan Smith, Cass R. Sunstein, Nassim Tabri, Joshua A. Tucker, Sander van der Linden, Paul van Lange, Kim A. Weeden, Michael J. A. Wohl, Jamil Zaki, Sean R. Zion, and Robb Willer (2020). "Using social and behavioural science to support COVID-19 pandemic response". *Nature Human Behaviour* 4.5, pp. 460–471.

Berger, Loıc, Nicolas Berger, Valentina Bosetti, Itzhak Gilboa, Lars Peter Hansen, Christopher Jarvis, Massimo Marinacci, and Richard D. Smith (2021). "Rational policymaking during a pandemic". *Proceedings of the National Academy of Sciences* 118.4, e2012704118.

Bogenschneider, Karen, Elizabeth Day, and Bret N. Bogenschneider (2021). "A window into youth and family policy: State policymaker views on polarization and research utilization". *The American Psychologist* 76.7, pp. 1143–1158.

Bol, Damien, Marco Giani, André Blais, and Peter John Loewen (2021). "The effect of COVID-19 lockdowns on political support: Some good news for democracy?" *European Journal of Political Research* 60.2, pp. 497–505.

Bornmann, Lutz, Robin Haunschild, and Rüdiger Mutz (2021). "Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases". *Humanities and Social Sciences Communications* 8.1, pp. 1–15.

Brainard, Jeffrey (2022). "Riding the twitter wave". *Science* 375.6587, pp. 1344–1347.

Bunea, Adriana, Raimondas Ibenskas, and Florian Weiler (2025). "Power, policy and national background: understanding interest groups-legislators ties on social media in the European Parliament". *Journal of European Public Policy*, pp. 1–27.

Cairney, Paul (2016). *The Politics of Evidence-Based Policy Making*. Springer.

Caplan, Nathan (1979). "The Two-Communities Theory and Knowledge Utilization". *American Behavioral Scientist* 22.3, pp. 459–470.

Caplan, Nathan, Russell J. Stambaugh, Andrea Morrison, and University of Michigan (1975). *The use of social science knowledge in policy decisions at the national level: a report to respondents*. Ann Arbor: Institute for Social Research, University of Michigan.

Castanho Silva, Bruno and Sven-Oliver Proksch (2022). "Politicians unleashed? Political communication on Twitter and in parliament in Western Europe". *Political Science Research and Methods* 10.4, pp. 776–792.

Côté, Isabelle M. and Emily S. Darling (2018). "Scientists on Twitter: Preaching to the choir or singing from the rooftops?" *FACETS* 3.1. Ed. by Stephen B. Heard, pp. 682–694.

Dodson, Elizabeth A., Nora A. Geary, and Ross C. Brownson (2015). "State legislators' sources and use of information: bridging the gap between research and policy". *Health Education Research* 30.6, pp. 840–848.

Dunbar, R. I. M., Valerio Arnaboldi, Marco Conti, and Andrea Passarella (2015). "The structure of online social networks mirrors those in the offline world". *Social Networks* 43, pp. 39–47.

Dunlop, Claire A. (2017). "The irony of epistemic learning: epistemic communities, policy learning and the case of Europe's hormones saga". *Policy and Society* 36.2, pp. 215–232.

Engel-Rebitzer, Eden, Daniel Camargo Stokes, Alison Buttenheim, Jonathan Purtle, and Zachary F. Meisel (2021). “Changes in legislator vaccine-engagement on Twitter before and after the arrival of the COVID-19 pandemic”. *Human Vaccines & Immunotherapeutics* 17.9, pp. 2868–2872.

Eysenbach, Gunther (2011). “Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact”. *Journal of Medical Internet Research* 13.4, e2012.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási (2018). “Science of science”. *Science* 359.6379, eaao0185.

Funk, Cary, Meg Hefferon, Brian Kennedy, and Courtney Johnson (2019). “Trust and Mistrust in Americans’ Views of Scientific Experts”.

Furnas, Alexander C., Timothy M. LaPira, and Dashun Wang (2025). “Partisan disparities in the use of science in policy”. *Science* 388.6745, pp. 362–367.

Gao, Jian, Yian Yin, Kyle R. Myers, Karim R. Lakhani, and Dashun Wang (2021). “Potentially long-lasting effects of the pandemic on scientists”. *Nature Communications* 12.1, p. 6188.

Garg, Prashant and Thiemo Fetzer (2025). “Political expression of academics on Twitter”. *Nature Human Behaviour*, pp. 1–18.

Gauchat, Gordon (2012). “Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010”. *American Sociological Review* 77.2, pp. 167–187.

Geddes, Marc (2021). “The webs of belief around ‘evidence’ in legislatures: The case of select committees in the UK House of Commons”. *Public Administration* 99.1, pp. 40–54.

Guntuku, Sharath Chandra, Jonathan Purtle, Zachary F Meisel, Raina M Merchant, and Anish Agarwal (2021). “Partisan Differences in Twitter Language Among US Legislators During the COVID-19 Pandemic: Cross-sectional Study”. *Journal of Medical Internet Research* 23.6, e27300.

Haas, Peter (1992). “Introduction: Epistemic Communities and International Policy Coordination”. *International Organization* 46, pp. 1–35.

Hassel, Anke and Kai Wegrich (2022). *How to Do Public Policy*. Oxford University Press.

He, Yuanmo and Milena Tsvetkova (2023). “A Method for Estimating Individual Socioeconomic Status of Twitter Users”. *Sociological Methods & Research*, p. 00491241231168665.

Heuer, Ruth, Peter Einaudi, and Kelly Kang (2023). *Research Doctorate Conferrals Rebound, Leading to Record Number of U.S. Doctorate Recipients in 2022*. Tech. rep. NSF 23-353. National Science Foundation.

Huberman, Michael (1994). “Research utilization: The state of the art”. *Knowledge and Policy* 7.4, pp. 13–33.

International Telecommunication Union (2021). *Measuring digital development - Facts and figures 2021*. Geneva: International Telecommunication Union.

Jones, Bryan D. and Frank R. Baumgartner (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.

Jungherr, Andreas (2016). “Twitter use in election campaigns: A systematic literature review”. *Journal of Information Technology & Politics* 13.1, pp. 72–91.

Kim, Taegyoon, Nitheesha Nakka, Ishita Gopal, Bruce A. Desmarais, Abigail Mancinelli, Jeffrey J. Harden, Hyein Ko, and Frederick J. Boehmke (2022). “Attention to the COVID-19 Pandemic on Twitter: Partisan Differences Among U.S. State Legislators”. *Legislative Studies Quarterly* 47.4, pp. 1023–1041.

Klar, Samara, Yanna Krupnikov, John Barry Ryan, Kathleen Searles, and Yotam Shmargad (2020). “Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work”. *PLOS ONE* 15.4, e0229446.

Lamberton, Cait and Andrew T. Stephen (2016). “A Thematic Exploration of Digital, Social Media, and Mobile Marketing: Research Evolution from 2000 to 2015 and an Agenda for Future Inquiry”. *Journal of Marketing* 80.6, pp. 146–172.

Lee, Nathan (2022). “Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers”. *American Political Science Review* 116.2, pp. 677–688.

Leininger, Arndt and Max Schaub (2023). “Strategic Alignment in Times of Crisis: Voting at the Dawn of a Global Pandemic”. *Political Behavior*.

Lewandowsky, Stephan and Klaus Oberauer (2016). “Motivated Rejection of Science”. *Current Directions in Psychological Science* 25.4, pp. 217–222.

Linden, Sander van der, Costas Panagopoulos, Flávio Azevedo, and John T. Jost (2021). "The Paranoid Style in American Politics Revisited: An Ideological Asymmetry in Conspiratorial Thinking". *Political Psychology* 42.1, pp. 23–51.

Mede, Niels G., Viktoria Cologna, Sebastian Berger, John Besley, Cameron Brick, Marina Joubert, Edward Maibach, Sabina Mihelj, Naomi Oreskes, and Mike S. Schäfer (2024). *Public communication about science across 68 countries: Global evidence on how people get information and communicate about science-related matters*.

Mede, Niels G. and Mike S. Schäfer (2020). "Science-related populism: Conceptualizing populist demands toward science". *Public Understanding of Science* 29.5, pp. 473–491.

Metaxas, Panagiotis, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O'Keefe, and Samantha Finn (2015). "What Do Retweets Indicate? Results from User Survey and Meta-Review of Research". *Proceedings of the International AAAI Conference on Web and Social Media* 9.1, pp. 658–661.

Mongeon, Philippe, Timothy D. Bowman, and Rodrigo Costas (2023). "An open data set of scholars on Twitter". *Quantitative Science Studies*, pp. 1–11.

National Research Council (2012). *Using Science as Evidence in Public Policy*. Washington, D.C.: National Academies Press.

Ouimet, Mathieu, Morgane Beaumier, Adrien Cloutier, Alexandre Côté, Éric Montigny, François Gélineau, Steve Jacob, and Stéphane Ratté (2023). "Use of research evidence in legislatures: a systematic review". *Evidence & Policy* -1.aop, pp. 1–18.

Parkhurst, Justin (2017). *The Politics of Evidence: From evidence-based policy to the good governance of evidence*. Taylor & Francis.

Peoples, Brandon K., Stephen R. Midway, Dana Sackett, Abigail Lynch, and Patrick B. Cooney (2016). "Twitter Predicts Citation Rates of Ecological Research". *PLOS ONE* 11.11, e0166570.

Priem, Jason, Heather Piwowar, and Richard Orr (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*.

Purtle, Jonathan, Elizabeth A. Dodson, Katherine Nelson, Zachary F. Meisel, and Ross C. Brownson (2018). "Legislators' Sources of Behavioral Health Research and Preferences for Dissemination: Variations by Political Party". *Psychiatric Services* 69.10, pp. 1105–1108.

Riphahn, Regina T. and Monika Schnitzer (2022). "Die Bedeutung wissenschaftlicher Evidenz für die Arbeit des Bundestags". *Wirtschaftsdienst* 102.4, pp. 270–272.

Roten, Fabienne Crettaz von (2004). "Gender Differences in Attitudes toward Science in Switzerland". *Public Understanding of Science* 13.2, pp. 191–199.

Rutjens, Bastiaan T., Steven J. Heine, Robbie M. Sutton, and Frenk Van Harreveld (2018). "Attitudes Towards Science". In: *Advances in Experimental Social Psychology*. Vol. 57. Elsevier, pp. 125–165.

Rutjens, Bastiaan T., Nikhil Sengupta, Romy van der Lee, Guido M. van Koningsbruggen, Jason P. Martens, André Rabelo, and Robbie M. Sutton (2022). "Science Skepticism Across 24 Countries". *Social Psychological and Personality Science* 13.1, pp. 102–117.

Rutjens, Bastiaan T., Robbie M. Sutton, and Romy van der Lee (2018). "Not All Skepticism Is Equal: Exploring the Ideological Antecedents of Science Acceptance and Rejection". *Personality and Social Psychology Bulletin* 44.3, pp. 384–405.

Seidel, Katja, Hans Verbeek, Sophie Fessel, and Florian Meer (2021). *Nutzen von wissenschaftlicher Evidenz: Erwartungen an wissenschaftliche Expertise : Bericht über eine Bundestagsbefragung im Rahmen der Leopoldina-Evidenzinitiative unter Leitung von Claudia Buch, Regina T. Riphahn und Monika Schnitzer*. Tech. rep. MyCoRe Community, vol. 27.

Senninger, Roman and Henrik B. Seeberg (2024). "Which Information Do Politicians Pay Attention To? Evidence from a Field Experiment and Interviews". *British Journal of Political Science*, pp. 1–18.

Spierings, Niels, Kristof Jacobs, and Nik Linders (2019). "Keeping an Eye on the People: Who Has Access to MPs on Twitter?" *Social Science Computer Review* 37.2, pp. 160–177.

Vivalt, Eva and Aidan Coville (2023). "How do policymakers update their beliefs?" *Journal of Development Economics* 165, p. 103121.

Walgrave, Stefaan and Yves Dejaeghere (2017). "Surviving Information Overload: How Elite Politicians Select Information". *Governance* 30.2, pp. 229–244.

Weiss, Carol H. (1979). "The Many Meanings of Research Utilization". *Public Administration Review* 39.5, pp. 426–431.

Weiss, Carol H. and Michael J. Bucuvalas (1980). "Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research". *American Sociological Review* 45.2, pp. 302–313.

Westerman, David, Patric R. Spence, and Brandon Van Der Heide (2014). "Social Media as Information Source: Recency of Updates and Credibility of Information". *Journal of Computer-Mediated Communication* 19.2, pp. 171–183.

Wojcieszak, Magdalena, Andreu Casas, Xudong Yu, Jonathan Nagler, and Joshua A. Tucker (2022). "Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity". *Science Advances* 8.39, eabn9418.

Yin, Yian, Jian Gao, Benjamin F Jones, and Dashun Wang (2021). "Coevolution of policy and science during the pandemic". *Science* 371.6525, pp. 128–130.

Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience

Online Appendix

Contents

Appendix A Data Collection, Sample, and Statistical Analyses

A.1 Data on legislators from 12 countries

The data were collected between November and December 2022. The legislator database centers around the Twitter profiles of elected legislators during 2022. The data collected extends across 12 countries with widespread Twitter usage. The countries in the study include the three largest democracies in both South and North America (Argentina, Brazil, Colombia, Mexico, Canada, and the U.S.), in addition to six European democracies (France, Italy, Ireland, Germany, Spain, and the U.K.)

The data collection strategy is based on the Twitter IDs and handles extracted from three sources: legislators' (i) Wikidata entries (entities P6552 for Twitter IDs and P2002 for user names), (ii) Google Knowledge Graph, and (iii) official legislatures' webpages. In cases in which the initial Twitter ID extraction did not result in matches for a legislator, I performed a manual searches on Twitter under the individual's name. In a final step, I validated the accounts retrieved to ensure that they belonged to the legislator.

The legislators on Twitter constitute 90% of all legislators across the 13 legislatures, ranging from 72% in the Italian Chamber of Deputies to 100% in the US Senate.

Using all retrieved account IDs, I collected legislators' Twitter data using their API via the rtweet package (Kearney, 2019a) between November and December 2022. I extracted the legislators':

- **Account core:** A general overview of their accounts, including the date of creation, username, location, number of followers, amongst others.
- **Following network:** A snapshot of all the accounts they followed on November 30, 2022.
- **Tweets:** The universe of their public tweets, retweets, and quotes since they joined the platform until December 5, 2022.
- **Likes:** The archive of liked posts by each account with an upper bound of the 3,200 most recent.

In addition to collecting legislator's relevant Twitter data, I extracted auxiliary legislator sociodemographic and political information. In a first step, I mimicked the procedure by the Comparative Legislators Database (CLD) (Göbel and Munzert, 2022) to extract relevant information using Wikidata and Wikipedia as a data source. These data include legislators' political party affiliation from the legislatures' Wikipedia lists and their dates of birth (Wikidata entity P569). I use these variables in the predictive models. In a further step, I identify whether legislators have research credentials by having been enrolled or completed second-stage tertiary education degrees (e.g., doctorates or equivalent). I gathered this information from the official legislature's websites, government open repositories of public servant curricula, legislative transparency portals managed by academic institutions (e.g., Congreso Visible by the Universidad de los Andes in Colombia), legislators' personal website biographies, and Wikipedia entries.

A.2 Data on researcher's social media profiles

The starting point for these data come from an open dataset of researchers on Twitter (Mongeon, Bowman, and Costas, 2023). The data builds on the Crossref Event Data release from January 2022 containing over 60 million Twitter posts with URL links to scholarly work. The authors extract the names of the authors associated to the works linked on the social media posts through the open repository of scholarly work, OpenAlex (Priem, Piwowar, and Orr, 2022). The connection between Twitter accounts to academic researchers relies on mapping the Twitter screen and user names onto the names of the authors associated to the papers mentioned in the posts. The authors outline a hierarchical matching process based on the name strings and variations of the handles and profile names. For instance, in one step matches are created, but not exclusively, after a full match between the authors name and the Twitter handle or user name (e.g., a paper whose author is Gregor Samsa is mentioned in a tweet by @gregorsamsa). In this dataset, Mongeon, Bowman, and Costas (2023) identify 423,920 Twitter accounts associated to authors of academic research studies with high precision.

Using these Twitter IDs, I collected researchers' core Twitter data using the Twitter REST API via the rtweet package (Kearney, 2019a) on December 12, 2022. When extracting the core features of these accounts, a total of 14,728 Twitter IDs did not relate to any active accounts.

Furthermore, I employed the OpenAlex author identifiers to extract auxiliary information about the individual researchers[1]. These data were retrieved as Author object calls from the OpenAlex API. The call returned last known institutional affiliations, total number of papers citing the author, and a list of concepts (inferred scientific fields) most frequently applied to research created by the author. Figure B1 and Table B3 present a descriptive overviews of the scientific disciplines and geographical associations of the identified researchers. The final researcher dataset contains core the Twitter and OpenAlex features of 409,192 identified authors of academic work.

Table B2 provides an overview of a selection of summary statistics from the legislators included in the study. Overall, the database contains Twitter, as well as sociodemographic and political information from 3,670 from a total set of 4,134 seating legislators (≈89%).

A.3 Measurement

Platform behaviors: I map the list of research producer accounts onto the legislators' friendship snapshots, tweets, retweets, quotes, and likes. I encode instances where there is a legislator-academic researcher account pairing. For the analyses, I employ the binary measures of matches, total counts of instances, and proportions of matches in reference to totals for all platform-specific behaviors. In other words, whether the legislators follow, mention, retweet, quote, and like academic researchers, how many times they do, and what share that represents of the overall historic legislator behavior.

Minimum possible following date: An important set of data points to understand the temporal development of legislator Twitter networks are the dates in which they start to follow other accounts. Nevertheless, the platform does not make these data available through their API. One way to gather this information is to extract snapshots of the networks in different periods and comparing the relationships between nodes. Still, given the API limits, these crawls can be time expensive and only allow for tracking forward-looking developments. That is to say, they can only showcase changes from the first snapshot. In this study, I use a different strategy relying on the default call to extract users' friendships returning a reverse chronologically ordered list. Using these data, I employ a method for inferring the creation time of connections between users in unidirectional networks using auxiliary information from the nodes (Meeder et al., 2011). The procedure involves mapping legislators' chronologically ordered edge list alongside the "record-breakers" (users with the latest account creation times) for each legislator. The follow time for a user is estimated to

[1] These data were collected through between December 15-23, 2022, corresponding to the December 2022 snapshot of OpenAlex

be the creation time of the most recent record-breaker among the users who that legislator follows.

Legislators' formal academic research background: I collect the highest educational qualification of legislators. I extract this information from the i) legislature's websites, b) governmental repositories of public servant curricula, iii) legislative transparency portals managed by academic institutions, iv) legislators' personal sites, and v) Wikidata entries. I map legislators' qualifications on the International Standard Classification of Education (ISCED 97) (Hoffmeyer-Zlotnik and Wolf, 2003). I identify a legislator to have research credentials when they have been enrolled or completed second-stage tertiary education degrees (e.g., doctorates or equivalent).

Scientific fields of academic researchers: I employ the concepts attached to the work of academic researchers from OpenAlex to extract their predicted scientific field. This measure is based on a hierarchical representation of scientific concepts. OpenAlex uses a slightly modified concept tree based on the one developed by Microsoft Academic Graph (Shen, Ma, and Wang, 2018). I use the top layer of the tree containing 19 concepts ranging from political to materials science. These classes are derived from the titles, abstracts, and hosts of the papers authored by the researcher. Further, I map the predicted fields into four branches i) humanities, ii) social, iii) natural, and iv) formal science (see *SI Appendix, Figure B1* for an overview of the distribution of research fields).

Party family: I utilize the party family encoding from the Manifesto Project. In the main analyses, I group electoral alliances, single issue, and ethnic parties onto the other category, as well as christian-democratic and conservative parties onto an umbrella conservative party category. In the cases where parties were not present in the dataset, I assigned them to a category employing the Manifesto Party Family Handbook (Lehmann et al., 2023). *SI Appendix, Table B9* presents the different political parties alongside their family classification.

Other covariates: For the analyses, I employ: Age divided in six categories (under 35, 35-44, 45-54, 55-64, and above 75), sex measured using a dummy variable (male=0; female=1), and the accounts' following and friends quartiles.

A.4 Statistical analyses

A.4.1 Models on predictors of engagement with researchers based on contextual and legislator characteristics

To explore predictors of engagement with researchers, I ran linear mixed-effects models with country of legislature random effects. I specified the models with the lmer() function from the lme4 package in R (Bates, Mächler, et al., 2015). I report both fixed-effects estimates, as well as conditional predictions. I compute the conditional predictions with age (under 35), party family (other), country region (Europe), system (presidential), and Q1 for followers and friends as references for categorical variables using the marginaleffects package in R (Arel-Bundock, 2023).

I specified models for a) the proportions of researchers-legislator instances in respect to totals for each behavior (see main body and section C.1) and b) the absolute number of researchers-legislator instances (see section C.2). The linear mixed-effect model is given by:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + ... + \beta_8 x_{8ij} + u_j + \epsilon_{ij}$$

where:

- y_{ij} is the proportion or absolute number of scholars for observation i in group j.
- $\beta_1, ..., \beta_8$ are the fixed effects coeffients for the covatiates.
- $u_j \sim N(0, \sigma_u^2)$ is the random intercept for each country j
- $\epsilon_{ij} \sim N(0, \sigma^2)$ is the residual error

A.4.2 Models on changes in the likelihood that an observed legislator Twitter event targets a researcher in the COVID compared to the pre-COVID period

To explore potential changes in the composition of the targets of Twitter events between the pre- and during-COVID periods, I estimated logistic mixed-effects models with country random effects. I specified the models with the gmler() function from the lme4 package in R (Bates, Mächler, et al., 2015). I compute the relative risks representing the ratio of the probability of an event in the COVID period to the probability of an outcome in a pre-COVID period from the log-odds using the avg_comparisons() function from the marginaleffects package in R (Arel-Bundock, 2023).

I specified models for behaviors targeting scholars in the list (pooled), as well as by scientific branch. The generalized linear mixed-effects model is given by:

$$\text{logit}(P(Y_i = 1)) = \beta_0 + \beta_1 X_{i1} + u_j$$

where:

- $\text{logit}(P(Y_i = 1))$ is the log-odds of individual i being a researcher or member of a specific research branch,
- β_0 is the fixed intercept,
- β_1 is the coefficient for the during COVID period X_{i1} (Declaration of COVID as an international public health emergency),
- $u_j \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept for the country of legislation, with variance σ_u^2.

Appendix B Supporting Figures and Tables

Table B1: Distribution of total and researcher-targeted legislator behaviors

Behavior	Overall						With academic researchers					
	No. of legs	Mean	Median	SD	Max	Min	No. of Legs[1]	Mean	Median	SD	Max	Min
Following	3670	1766.5	990.0	3865.2	127189	1	3151 (85.9%)	29.1	10	56.7	1146	1
Mentioning	3601	1102.4	562.0	1588.6	21726	1	2272 (63.1%)	13.2	5	34.3	800	1
Retweeting	3619	3885.3	1476.0	8048.5	200781	1	2658 (73.4%)	62.2	11	211.3	4901	1
Quote tweeting	3420	557.9	239.0	1036.9	23402	1	1880 (55%)	12.9	4	36.4	940	1
Replying	3589	1317.6	370.0	3318.0	65903	1	2017 (56.2%)	36.0	6	203.6	7980	1
Liking	3580	1808.7	1823.5	1254.5	3262	1	2828 (79%)	31.7	12	55.4	726	1

[1] The percentage represents the share of legislators in relation to the number that engage in such behavior. For instance, 86% of the 3670 legislators that follow accounts on Twitter have an academic researcher match.

Table B2: Summary statistics of legislators included in the study

				Party family							Twitter Friends			Twitter Followers			Age		
	No. Legs	Research degree	Female	CON	ECO	LIB	NAT	SD	LEF	Other	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
Argentina	61	3 (4.9%)	26 (42.6%)	7 (11.5%)	0 (0%)	0 (0%)	0 (0%)	17 (27.9%)	0 (0%)	37 (60.7%)	12	7276	851.1	428	1177991	42331.1	34	79	56.2
Brazil	76	5 (6.6%)	10 (13.2%)	13 (17.1%)	1 (1.3%)	10 (13.2%)	15 (19.7%)	12 (15.8%)	9 (11.8%)	16 (21.1%)	1	19936	1487.5	247	4025807	314161.7	39	80	59.0
Canada	324	14 (4.3%)	101 (31.2%)	111 (34.3%)	2 (0.6%)	155 (47.8%)	0 (0%)	25 (7.7%)	0 (0%)	31 (9.6%)	1	26036	1739.1	181	6357427	39611.9	24	81	52.6
Colombia	104	3 (2.9%)	32 (30.8%)	27 (26%)	9 (8.7%)	23 (22.1%)	4 (3.8%)	0 (0%)	19 (18.3%)	22 (21.2%)	27	24290	1890.4	168	1780086	113715.2	31	76	51.2
France	548	29 (5.3%)	205 (37.4%)	84 (15.3%)	23 (4.2%)	209 (38.1%)	88 (16.1%)	30 (5.5%)	91 (16.6%)	23 (4.2%)	1	9676	1359.1	0	2894718	21771.4	22	79	48.8
Germany	583	95 (16.3%)	207 (35.5%)	127 (21.8%)	110 (18.9%)	90 (15.4%)	64 (11%)	153 (26.2%)	35 (6%)	4 (0.7%)	1	8425	984.7	5	1072411	19399.4	24	74	47.3
Ireland	154	4 (2.6%)	35 (22.7%)	73 (47.4%)	12 (7.8%)	0 (0%)	0 (0%)	12 (7.8%)	5 (3.2%)	52 (33.8%)	40	7493	1970.7	528	449917	20152.3	25	77	51.1
Italy	286	18 (6.3%)	85 (29.7%)	94 (32.9%)	11 (3.8%)	0 (0%)	52 (18.2%)	62 (21.7%)	0 (0%)	67 (23.4%)	2	14081	993.5	0	1754299	36143.1	25	76	48.8
Mexico	114	20 (17.5%)	61 (53.5%)	21 (18.4%)	5 (4.4%)	0 (0%)	0 (0%)	26 (22.8%)	59 (51.8%)	3 (2.6%)	12	39006	2242.8	237	3212451	117791.6	32	92	55.0
Spain	314	40 (12.7%)	133 (42.4%)	79 (25.2%)	0 (0%)	10 (3.2%)	44 (14%)	106 (33.8%)	2 (0.6%)	73 (23.2%)	1	98546	2452.5	66	1717578	48252.2	26	75	51.9
UK	580	24 (4.1%)	209 (36%)	305 (52.6%)	1 (0.2%)	14 (2.4%)	0 (0%)	196 (33.8%)	7 (1.2%)	57 (9.8%)	1	26769	2356.4	2	4776780	64919.6	26	82	52.0
US House	426	27 (6.3%)	121 (28.4%)	208 (48.8%)	0 (0%)	0 (0%)	0 (0%)	218 (51.2%)	0 (0%)	0 (0%)	1	47205	1983.3	316	13464680	168925.0	27	86	59.1
US Senate	100	5 (5%)	24 (24%)	50 (50%)	0 (0%)	0 (0%)	0 (0%)	48 (48%)	0 (0%)	2 (2%)	1	127189	4142.8	6838	12510878	755868.3	35	89	65.2
Pooled	3670	287 (7.8%)	1249 (34%)	1199 (32.7%)	174 (4.7%)	511 (13.9%)	267 (7.3%)	905 (24.7%)	227 (6.2%)	387 (10.5%)	1	127189	1766.5	0	13464680	82174.9	22	92	52.0

Figure B1: Overview of academic producers in the (Mongeon, Bowman, and Costas, 2023) list

Predicted field	Twitter accounts	Followed by LEGs
Pooled	**409192**	27295 (6.7%)
Political science (Social science)	39172	7418 (18.9%)
Economics (Social science)	8389	1404 (16.7%)
Art (Humanities)	2054	317 (15.4%)
History (Humanities)	3373	475 (14.1%)
Philosophy (Humanities)	6179	820 (13.3%)
Sociology (Social science)	4292	541 (12.6%)

Predicted field	Twitter accounts	Followed by LEGs
Business (Applied science)	3690	362 (9.8%)
Geography (Natural science)	4241	276 (6.5%)
Psychology (Social science)	56815	3524 (6.2%)
Computer science (Formal science)	29013	1527 (5.3%)
Environmental science (Natural science)	6493	345 (5.3%)
Geology (Natural science)	5005	257 (5.1%)
Engineering (Formal science)	3346	167 (5%)

Predicted field	Twitter accounts	Followed by LEGs
Medicine (Applied science)	94718	4439 (4.7%)
Mathematics (Formal science)	7482	325 (4.3%)
Physics (Natural science)	15217	555 (3.6%)
Biology (Natural science)	83120	2814 (3.4%)
Chemistry (Natural science)	12176	255 (2.1%)
Materials science (Natural science)	2759	56 (2%)
Not categorized (No category)	21658	1418 (6.5%)

Table B3: Geographical distribution of researchers based on their last known institutional affiliation

Country	No. of acc	Country	No. of acc	Country	No. of acc	Country	No. of acc	Country	No. of acc
United States	115101	Indonesia	621	Costa Rica	99	Namibia	20	Nicaragua	7
United Kingdom	59562	Egypt	608	Iraq	97	Burundi	19	Uzbekistan	7
No info	42970	Nigeria	585	Tunisia	96	Mali	19	Angola	6
Australia	19990	Bangladesh	529	Lithuania	95	Albania	18	Gibraltar	6
Canada	18589	Ecuador	508	Kuwait	93	Congo - Brazzaville	18	Kyrgyzstan	6
Spain	15604	Malaysia	504	Bulgaria	92	Mongolia	18	Liechtenstein	6
Germany	15362	Taiwan	504	Malawi	89	Yemen	18	Papua New Guinea	6
France	8825	Philippines	500	Oman	87	St. Kitts & Nevis	17	Somalia	6
Netherlands	8614	Kenya	496	North Macedonia	79	Burkina Faso	16	Tajikistan	6
India	8490	Lebanon	425	Rwanda	76	Réunion	16	Eswatini	5
Italy	8375	Peru	383	Ukraine	75	Brunei	15	Guinea	5
Brazil	5867	United Arab Emirates	365	Algeria	70	Greenland	15	Monaco	5
Switzerland	5527	Hungary	350	Zambia	59	Libya	15	Seychelles	5
Sweden	4826	Slovenia	350	Cameroon	58	Puerto Rico	15	Andorra	4
Ireland	4510	Uruguay	307	Hong Kong SAR China	58	Benin	13	Bhutan	4
Finland	4125	Ghana	294	Bahrain	55	Barbados	12	French Polynesia	4
Belgium	4070	Thailand	286	Kazakhstan	55	Grenada	12	Gabon	4
Denmark	3937	Nepal	282	Cuba	54	Guadeloupe	12	Samoa	4
China	3927	Qatar	275	Bolivi		Madagascar	12	Belize	3
Japan	3597	Romania	262	Cambodia	51	Montenegro	12	Bermuda	3
Norway	3293	Uganda	248	Palestinian Territories	48	Mozambique	12	Curaçao	3
Turkey	2662	Cyprus	226	Guatemala	45	New Caledonia	12	Falkland Islands	3
Mexico	2361	Luxembourg	222	Paraguay	45	Afghanistan	11	Palau	3
Portugal	2190	Croatia	216	Bosnia & Herzegovina	37	Azerbaijan	11	Cayman Islands	2
Austria	2115	Ethiopia	213	Senegal	35	El Salvador	11	Chad	2
South Africa	2102	Estonia	209	Moldova	32	Haiti	11	Guinea-Bissau	2
New Zealand	2037	Georgia	207	Sudan	32	Mauritius	11	Guyana	2
Israel	2029	Serbia	183	Armenia	31	Sierra Leone	11	Lesotho	2
Chile	1631	Iceland	171	Belarus	31	South Sudan	11	North Korea	2
Argentina	1630	Venezuela	163	Jamaica	31	Togo	11	Turkmenistan	2
Saudi Arabia	1509	Tanzania	155	Syria	31	Laos	10	Aruba	1
Colombia	1476	Morocco	144	Botswana	29	Niger	10	Cape Verde	1
Poland	1339	Sri Lanka	143	Fiji	29	Svalbard & Jan Mayen	10	Isle of Man	1
Singapore	1113	Slovakia	141	Gambia	27	Bahamas	9	Liberia	1
Greece	1034	Jordan	118	Côte d'Ivoire	25	Faroe Islands	9	Micronesia (Federated States of)	1
South Korea	1004	Vietnam	114	Trinidad & Tobago	25	Myanmar (Burma)	9	Montserrat	1
Pakistan	990	Panama	105	Antigua & Barbuda	23	French Guiana	8	San Marino	1
Czechia	902	Latvia	102	Congo - Kinshasa	23	Jersey	7	St. Lucia	1
Russia	738	Malta	102	Dominican Republic	22	Maldives	7	Suriname	1
Iran	703	Zimbabwe	101	Honduras	20				

Table B5: Academic researchers with the most retweets by legislators following the declaration of COVID

Twitter handle	Name	Scientific branch	Last known affiliation	# pre	# post	# legs
@G_Caballero_M	Gonzalo Caballero-Miguez	Social science	Universidade de Vigo	0	261	11
@TorstenBell	Torsten Bell	Social science	—	0	189	53
@DrTedros	Tedros Adhanom Ghebreyesus	Applied science	World Health Organization	0	109	78
@jasonleitch	J. Leitch	Social science	Scottish Government	0	108	28
@josepcosta	Josep Costa	Social science	—	0	105	8
@perezlozano	Lluis Perez-Lozano	Social science	Pompeu Fabra University	0	88	10
@cblackst	Cindy Blackstock	Social science	McGill University	0	81	12
@premnsikka	Prem Sikka	Applied science	University of Essex	0	74	9
@janephilpott	Jane Philpott	Applied science	University of Toronto	0	74	18
@Orla_Hegarty	Orla Hegarty	Social science	University College Dublin	0	69	12
@martamartirio	Marta Martín-Llaguno	Social science	University of Alicante	0	69	11
@c_drosten	Christian Drosten	Natural science	Humboldt-Universität zu Berlin	0	68	43
@XSalaimartin	Xavier Sala-i-Martin	Social science	MBIA	0	66	6
@juanrallo	Juan Ramón Rallo	Humanities	IE University	0	63	28
@RAWnGreen	Julia K. Green	Natural science	University of California, Berkeley	17	63	1
@uksciencechief	Patrick Vallance	Applied science	GlaxoSmithKline	0	62	40
@doctor_oxford	Rachel Clarke	Applied science	University of Oxford	0	56	30
@RZitelmann	Rainer Zitelmann	Social science	National Coalition of Independent Scholars	0	55	6
@ronan_glynn	Robert J. Glynn	Applied science	Brigham and Women's Hospital	0	55	18
@ZulmaCucunuba	Zulma M. Cucunubá	Applied science	Imperial College London	0	52	4
@JaimePalomera	Jaime Palomera	Social science	University of Barcelona	0	51	13
@GabrielScally	Gabriel Scally	Applied science	University of Bristol	0	49	22
@schnellenbachj	Jan Schnellenbach	Social science	Brandenburg University of Technology	0	49	3
@gebelque	Germà Bel	Social science	University of Barcelona	0	44	5
@jaumepadros	J. Padrós-Selma	Social science	West Virginia University College of Law	0	44	11

Table B6: Academic researchers with the most mentions by legislators following the declaration of COVID

Twitter handle	Name	Scientific branch	Last known affiliation	# pre	# post	# legs
@G_Caballero_M	Gonzalo Caballero-Miguez	Social science	Universidade de Vigo	0	52	9
@DrTedros	Tedros Adhanom Ghebreyesus	Applied science	World Health Organization	0	25	13
@huw4ogmore	Harriet Harden-Davies	Social science	University of Wollongong	1	20	1
@CDCDirector	Rochelle P Walensky	Social science	Centers for Disease Control and Prevention	0	19	15
@JoMalagon	Jonathan Malagon	No category	—	0	19	11
@c_drosten	Christian Drosten	Natural science	Humboldt-Universität zu Berlin	0	14	14
@SteveFDA	Stephen M. Hahn	Applied science	Annenberg Public Policy Center	0	14	10
@Martin_M_Guzman	Martin Guzman	Social science	Columbia University	0	12	8
@ilariacapua	Ilaria Capua	Applied science	University of Florida Health	0	11	5
@samuel_garcias	—	No category	—	0	10	8
@ASlavitt	Andrew Slavitt	Applied science	Centers for Medicare and Medicaid Services	0	8	3
@jasonleitch	J. Leitch	Social science	Scottish Government	0	8	6
@uksciencechief	Patrick Vallance	Applied science	GlaxoSmithKline	0	8	8
@ashishkjha	Ashish K. Jha	Applied science	Brown University	0	7	1
@JanezPotocnik22	Janez Potočnik	Social science	United Nations Environment Programme	0	6	1
@DrJV75	Justin Varney	Applied science	Public Health England	3	6	1
@hendrikstreeck	—	No category	—	0	6	6
@gabriel_zucman	Gabriel Zucman	Social science	University of California, Berkeley	0	6	3
@WRicciardi	—	No category	—	0	6	4
@Healthmac	Christopher Mackie	Applied science	Middlesex London Health Unit	0	6	3
@AlanDersh	Alan M. Dershowitz	Social science	—	0	6	2
@CarloMasala1	Carlo Masala	Social science	Bundeswehr University Munich	0	6	3
@GaviSeth	Seth Berkley	Applied science	Gavi	0	6	4
@lugaricano	Luis Garicano	Social science	IE University	0	5	4
@MarvinJRees	Marvin Rees	Social science	—	0	5	4

Table B7: Academic researchers with the most quotes by legislators following the declaration of COVID

Twitter handle	Name	Scientific branch	Last known affiliation	# pre	# post	# legs
@TorstenBell	Torsten Bell	Social science	—	0	33	31
@janephilpott	Jane Philpott	Applied science	University of Toronto	0	23	17
@c_drosten	Christian Drosten	Natural science	Humboldt-Universität zu Berlin	0	22	10
@DrTedros	Tedros Adhanom Ghebreyesus	Applied science	World Health Organization	0	20	18
@G_Caballero_M	Gonzalo Caballero-Miguez	Social science	Universidade de Vigo	8	17	2
@hans_kluge	Hans Kluge	Applied science	World Health Organization Regional Office for Europe	0	14	13
@ASlavitt	Andrew Slavitt	Applied science	Centers for Medicare and Medicaid Services	0	13	10
@juanrallo	Juan Ramón Rallo	Humanities	IE University	0	11	8
@Orla_Hegarty	Orla Hegarty	Social science	University College Dublin	0	10	4
@cblackst	Cindy Blackstock	Social science	McGill University	9	10	3
@jdportes	Jonathan Portes	Social science	King's College London	0	9	5
@doctor_oxford	Rachel Clarke	Applied science	University of Oxford	7	9	6
@AlanDersh	Alan M. Dershowitz	Social science	—	0	8	4
@ronan_glynn	Robert J. Glynn	Applied science	Brigham and Women's Hospital	0	8	5
@ianbremmer	Ian A. (Ian Arthur) Bremmer	Social science	Hoover Institution	0	8	8
@MacaesBruno	Bruno Macaes	Social science	—	0	7	2
@GabrielScally	Gabriel Scally	Applied science	University of Bristol	0	7	4
@uksciencechief	Patrick Vallance	Applied science	GlaxoSmithKline	0	7	6
@nntaleb	Nassim Nicholas Taleb	Formal science	New York University	0	7	3
@oriolmitja	Oriol Mitjà	Applied science	Fight AIDS Foundation	0	7	3
@Miguel__Lorente	Miguel Lorente-Acosta	Humanities	University of Granada	5	7	1
@MaxCRoser	Max Roser	Social science	Center for Global Development	0	7	5
@jasonleitch	J. Leitch	Social science	Scottish Government	0	6	6
@premnsikka	Prem Sikka	Applied science	University of Essex	0	6	6
@jsuedekum	Jens Suedekum	Social science	Heinrich Heine University Düsseldorf	0	6	5

Table B8: Academic researchers with the most replies by legislators following the declaration of COVID

Twitter handle	Name	Scientific branch	Last known affiliation	# pre	# post	# legs
@CarloMasala1	Carlo Masala	Social science	Bundeswehr University Munich	36	69	10
@BachmannRudi	Ruediger Bachmann	Social science	University of Notre Dame	0	33	4
@SDullien	Sebastian Dullien	Social science	Hans Böckler Foundation	0	29	6
@mquijoux	Maxime Quijoux	Social science	Laboratoire Interdisciplinaire pour la Sociologie Economique	6	21	1
@StephanieCarvin	Stephanie Carvin	Social science	University of Ottawa	0	21	5
@thesismum	Anja Katharina Peters	Social science	—	0	17	4
@AnMailleach	Eoin O'Malley	Social science	Dublin City University	14	16	9
@AchimTruger	Achim Truger	Social science	University of Duisburg-Essen	0	16	5
@jsuedekum	Jens Suedekum	Social science	Heinrich Heine University Düsseldorf	0	15	3
@m_kubiciel	Michael Kubiciel	Social science	University of Cologne	0	15	4
@DrGrandMal	Dennis Müller	Natural science	Heidelberg University	0	14	5
@HerrLuehmann	Michael Lühmann	Natural science	—	0	13	4
@Lars_Feld	Lars P. Feld	Social science	Walter Eucken Institut	0	13	7
@gavinjdaly	Gavin Daly	Social science	National University of Ireland, Maynooth	0	11	4
@tatterededge	Lelainia Lloyd	Social science	—	4	11	1
@PolProfSteve	Steven Fielding	Social science	University of Nottingham	0	11	7
@lofferg	Christopher Gohl	Social science	Leibniz-Institut für Wissensmedien	0	10	3
@JuergenZimmerer	Jürgen Zimmerer	Social science	University of Sheffield	2	10	1
@wargonm	Mathias Wargon	Applied science	Centre Hospitalier Saint-Denis	0	10	5
@devisridhar	Devi Sridhar	Applied science	University of Edinburgh	0	9	2
@DannyFiler	Danny Filer	Applied science	University College London	0	9	3
@cataperezcorrea	Catalina Pérez Correa	Social science	University of Iceland	1	9	1
@Puettmann_Bonn	Andreas Püttmann	Social science	—	0	9	4
@drcrouchback	Paul W Keeley	Applied science	University of Glasgow	1	9	1
@T_Ortelt	Tobias R. Ortelt	Formal science	TU Dortmund University	0	9	4

Table B9: Political parties and families by country

Country	Party	Family	Country	Party	Family
Argentina	Frente PRO	CON	France	DVG	SOC
Argentina	Unión Cívica Radical	SOC	France	POI	LEF
Brazil	MDB	SOC	France	Horizons-CCB	CON
Brazil	UNIÃO	DIV	France	LR-Nouvelle énergie	CON
Brazil	PL	NAT	France	DVG	ETH
Brazil	PSD	LIB	France	UDI	CON
Brazil	REDE (PSOL REDE)	ECO	France	Tavini	LEF
Brazil	PSDB (PSDB Cidadania)	CHR	France	LREM-PE	LIB
Brazil	PT (FE Brasil)	LEF	France	LREM-GNC	LIB
Brazil	PODE	SIP	France	LREM, Cap21	LIB
Brazil	Republicanos	CON	France	GUSR	ETH
Brazil	PDT	SOC	France	PRG	MI
Brazil	Cidadania (PSDB Cidadania)	CHR	France	LREM - TdP	LIB
Brazil	PSB	SOC	France	PNC	ETH
Brazil	PP	CON	France	REG	ETH
Brazil	PSC	NAT	France	PS	SOC
Brazil	PROS	DIV	France	DVC	LIB
Canada	Liberal	LIB	France	LREM, EC	LIB
Canada	New Democratic	SOC	France	DVD	MI
Canada	Green	ECO	Germany	CDU/CSU (CDU)	CHR
Canada	Conservative	CON	Germany	Grüne	ECO
Canada	Bloc Québécois	SIP	Germany	FDP	LIB
Canada	Independent	MI	Germany	SPD	SOC
Colombia	CD	CON	Germany	Linke	LEF
Colombia	MAIS	ETH	Germany	CDU/CSU (CSU)	CHR
Colombia	AV	ECO	Germany	AfD	NAT
Colombia	PUG	SIP	Germany	Fraktionslos	MI
Colombia	CH	LEF	Ireland	Fianna Fáil	CON
Colombia	CR	LIB	Ireland	Fine Gael	CHR
Colombia	MAIS	ETH	Ireland	Sinn Féin	SIP
Colombia	UP	LEF	Ireland	Green Party	ECO
Colombia	PLC	LIB	Ireland	Independent	MI
Colombia	PDA	LEF	Ireland	Solidarity–People Before Profit	LEF
Colombia	MIRA	NAT	Ireland	Labour	SOC
Colombia	AV	ECO	Ireland	Aontú	CHR
Colombia	ADA	ETH	Ireland	Social Democrats	SOC
Colombia	ASI	ETH	Ireland	Independents 4 Change	DIV
Colombia	COM	LEF	Italy	PARTITO DEMOCRATICO - ITALIA DEMOCRATICA E PROGRESSISTA	SOC
Colombia	VO	ECO	Italy	FRATELLI D'ITALIA	CON
Colombia	PC	CON	Italy	LEGA - SALVINI PREMIER	NAT
Colombia	ASI	ETH	Italy	AZIONE - ITALIA VIVA - RENEW EUROPE	DIV
Colombia	CJL	NAT	Italy	FORZA ITALIA - BERLUSCONI PRESIDENTE - PPE	CON
Colombia	AICO	ETH	Italy	MISTO-+EUROPA	SIP
France	LREM	LIB	Italy	MOVIMENTO 5 STELLE	SIP
France	G.s	ECO	Italy	NOI MODERATI (NOI CON L'ITALIA, CORAGGIO ITALIA, UDC, ITALIA AL CENTRO)-MAIE	DIV
France	TdP	LIB	Italy	ALLEANZA VERDI E SINISTRA	ECO
France	DVD	LIB	Italy	MISTO-MINORANZE LINGUISTICHE	ETH
France	LR	CON	Mexico	PAN	CON
France	RN	NAT	Mexico	Morena	LEF
France	MoDem	LIB	Mexico	Sin Partido	MI
France	LFI	LEF	Mexico	PRI	SOC
France	DVC	LIB	Mexico	Partido Movimiento Ciudadano	SOC
France	EELV	ECO	Mexico	Partido Encuentro Social	CON
France	Horizons	CON	Mexico	PRD	SOC
France	PS	SOC	Mexico	Partido Verde	ECO
France	PCF	LEF	Mexico	Worker's Party	LEF
France	LREM-TdP	LIB	Mexico	Movimiento Ciudadano	SOC
France	LND	ECO	Spain	PSOE	SOC
France	PRV	LIB	Spain	JxCat-JUNTS (Junts)	ETH
France	LREM-EC	LIB	Spain	EC-UP	ETH
France	PRV	ETH	Spain	PSC-PSOE	SOC
France	Agir	LIB	Spain	Vox	NAT
France	DVC	ETH	Spain	ERC-S	ETH
France	FP	LIB	Spain	PSE-EE-PSOE	SOC
France	DVG	MI	Spain	PsdeG-PSOE	SOC
France	LC	ETH	Spain	EH Bildu	ETH
France	LFI-E!	LEF	Spain	UP	ETH
France	REV	LEF	Spain	PP	CON
France	FaC	ETH	Spain	ECP-GUAYEM EL CANVI	ETH
France	MDES	LEF	Spain	MÁS PAÍS-EQUO	LEF
France	EÉLV	ECO	Spain	NA+	ETH
France	PPDG	SOC	Spain	EAJ-PNV	ETH
France	LREM	LIB	Spain	JxCat-JUNTS (PDeCAT)	ETH

Table B9: Political parties and families by country

Country	Party	Family	Country	Party	Family
France	Agir	CON	Spain	Cs	LIB
France	LR	CON	Spain	CUP-PR	ETH
France	DVG	LEF	Spain	MÉS COMPROMÍS	ETH
France	GE	ECO	Spain	BNG	ETH
France	DVD	CON	Spain	CCa-PNC-NC	ETH
France	CE	LIB	Spain	UP	ETH
France	DLF	MI	Spain	Cs	LIB
France	LR-A droite !	CON	Spain	PP-FORO	CON
France	RR	LIB	Spain	NC-CCa-PNC	ETH
France	FGPS	SOC	Spain	PP-FORO	CON
France	EXD	MI	Spain	PRC	ETH
France	LFI-PG	LEF	UK	Labour Co-operative	SOC
France	LR-SL	CON	UK	Conservative	CON
France	DVD	LIB	UK	Labour	SOC
France	DVD	ETH	UK	Liberal Democrats	LIB
France	PRV	LIB	UK	Sinn Féin	LEF
France	LS	NAT	UK	Scottish National	ETH
France	LFI-PD	LEF	UK	Green	ECO
France	LREM-RSM	LIB	UK	SDLP	SOC
France	PLR	LEF	UK	Plaid Cymru	ETH
France	UDI	ETH	UK	DUP	ETH
France	Horizons	LIB	UK	Alliance	LIB
France	Horizons-LREM	CON	UK	Speaker	CON
France	LREM	CON	UK	SNP	ETH
France	LFI-RÉ974	LEF	US	Democratic	SOC
France	PRV-LREM	LIB	US	Republican	CON
France	PPM	SOC	US	Independent	MI

Table B10: Estimated effects of legislator and legislature characteristics on the proportion of researchers across behaviors. (*Fig. 3* in main body and *Figs. C1-C5*)

	Following	Mentioning	Retweeting	Quote tweeting	Replying	Liking
Doctoral studies	0.009***	0.007***	0.006***	0.006***	0.008***	0.007***
	[0.008, 0.011]	[0.005, 0.008]	[0.005, 0.008]	[0.004, 0.008]	[0.005, 0.011]	[0.005, 0.010]
Female	0.000	0.000	−0.001	−0.001	0.001	0.000
	[−0.001, 0.001]	[−0.001, 0.001]	[−0.002, 0.000]	[−0.002, 0.001]	[−0.001, 0.003]	[−0.002, 0.001]
Age: 35-44 (ref: Under 35)	0.000	0.001	−0.001	0.001	0.001	0.001
	[−0.002, 0.001]	[−0.001, 0.003]	[−0.003, 0.001]	[−0.001, 0.004]	[−0.002, 0.005]	[−0.002, 0.004]
45-54	−0.003**	0.000	−0.003**	0.000	0.001	−0.001
	[−0.005, −0.001]	[−0.002, 0.001]	[−0.005, −0.001]	[−0.002, 0.003]	[−0.003, 0.004]	[−0.004, 0.002]
55-64	−0.003**	0.000	−0.002+	0.002	0.000	0.001
	[−0.004, −0.001]	[−0.002, 0.002]	[−0.004, 0.000]	[−0.001, 0.004]	[−0.003, 0.003]	[−0.002, 0.003]
65-74	−0.004***	0.000	−0.003**	0.001	−0.001	−0.001
	[−0.006, −0.002]	[−0.002, 0.002]	[−0.005, −0.001]	[−0.002, 0.003]	[−0.005, 0.003]	[−0.004, 0.002]
above 75	−0.005**	−0.001	−0.004*	0.000	−0.002	−0.003
	[−0.008, −0.002]	[−0.004, 0.002]	[−0.008, −0.001]	[−0.004, 0.004]	[−0.007, 0.003]	[−0.008, 0.002]
Party family: Conservative (ref: Other)	−0.003***	−0.001	−0.003**	−0.003*	−0.003+	−0.003*
	[−0.005, −0.002]	[−0.002, 0.001]	[−0.004, −0.001]	[−0.005, 0.000]	[−0.006, 0.000]	[−0.006, −0.001]
Ecological	0.008***	0.002	0.004**	0.002	0.003	0.004*
	[0.005, 0.010]	[−0.001, 0.004]	[0.001, 0.007]	[−0.002, 0.005]	[−0.001, 0.008]	[0.000, 0.008]
Liberal	0.001	0.000	−0.001	−0.002	0.004*	0.002
	[−0.001, 0.003]	[−0.002, 0.002]	[−0.003, 0.001]	[−0.004, 0.001]	[0.000, 0.007]	[−0.001, 0.005]
Nationalistic and radical right	−0.005***	−0.002	−0.001	−0.003+	−0.006**	−0.005**
	[−0.007, −0.003]	[−0.004, 0.001]	[−0.004, 0.001]	[−0.006, 0.000]	[−0.009, −0.002]	[−0.008, −0.001]
Social democratic	0.003**	0.002+	0.001	0.001	0.000	0.002
	[0.001, 0.004]	[0.000, 0.003]	[−0.001, 0.003]	[−0.002, 0.003]	[−0.003, 0.003]	[−0.001, 0.005]
Socialist and other left	0.000	−0.001	0.000	−0.003+	−0.001	0.000
	[−0.002, 0.003]	[−0.003, 0.001]	[−0.003, 0.002]	[−0.006, 0.000]	[−0.005, 0.003]	[−0.004, 0.004]
Followers qtl: Q2 (ref: Q1)	0.000	0.001	0.000	0.001	−0.003*	−0.001
	[−0.001, 0.002]	[−0.001, 0.002]	[−0.001, 0.002]	[0.000, 0.003]	[−0.005, 0.000]	[−0.003, 0.001]
Q3	0.004***	0.003***	0.003***	0.003**	0.000	0.004***
	[0.002, 0.005]	[0.002, 0.005]	[0.002, 0.005]	[0.001, 0.005]	[−0.003, 0.002]	[0.002, 0.007]
Q4	0.007***	0.005***	0.003***	0.005***	0.000	0.006***
	[0.005, 0.008]	[0.003, 0.006]	[0.002, 0.005]	[0.003, 0.007]	[−0.003, 0.002]	[0.003, 0.008]
Friends qtl: Q2 (ref: Q1)	0.001*	0.000	0.000	−0.002*	0.001	0.000
	[0.000, 0.003]	[−0.001, 0.002]	[−0.002, 0.001]	[−0.004, 0.000]	[−0.001, 0.003]	[−0.002, 0.002]
Q3	0.002***	0.001	0.000	−0.001	0.002*	0.001
	[0.001, 0.004]	[0.000, 0.002]	[−0.001, 0.002]	[−0.003, 0.001]	[0.000, 0.005]	[−0.001, 0.004]
Q4	0.003***	0.001	0.002*	0.000	0.003*	0.003**
	[0.001, 0.004]	[−0.001, 0.002]	[0.000, 0.003]	[−0.002, 0.002]	[0.000, 0.005]	[0.001, 0.006]
Region: LAC (ref: EUR)	−0.002	0.001	−0.002	−0.001	−0.003	−0.003
	[−0.009, 0.005]	[−0.004, 0.006]	[−0.009, 0.004]	[−0.007, 0.006]	[−0.010, 0.003]	[−0.012, 0.006]
NAm	−0.002	−0.001	0.000	0.000	−0.002	0.000
	[−0.009, 0.004]	[−0.005, 0.004]	[−0.005, 0.006]	[−0.006, 0.005]	[−0.008, 0.003]	[−0.008, 0.008]
Parliamentary republic	0.010**	0.005*	0.007*	0.007*	0.008**	0.010*
	[0.003, 0.016]	[0.000, 0.009]	[0.001, 0.012]	[0.001, 0.012]	[0.003, 0.013]	[0.002, 0.018]
SD (Observations)	0.013	0.013	0.014	0.017	0.022	0.019
Num.Obs.	3381	3318	3335	3145	3311	3295
R2 Marg.	0.252	0.082	0.109	0.068	0.075	0.117
R2 Cond.	0.308	0.119	0.155	0.101	0.092	0.161
ICC	0.1	0.0	0.1	0.0	0.0	0.0
RMSE	0.01	0.01	0.01	0.02	0.02	0.02

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

Table B11: Pooled scholars. Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5a* in main body)

	Following	Mentioning	Retweeting	Quote tweeting	Replying
Post-COVID	0.343^{***}	0.306^{***}	0.162^{***}	0.185^{***}	-0.052
	[0.296, 0.390]	[0.202, 0.411]	[0.126, 0.198]	[0.096, 0.274]	[−0.115, 0.011]
Num.Obs.	334 670	315 542	925 499	166 371	266 846
R2 Marg.	0.008	0.007	0.002	0.002	0.000
R2 Cond.	0.097	0.054	0.145	0.075	0.256
ICC	0.1	0.0	0.1	0.1	0.3
RMSE	0.15	0.07	0.11	0.11	0.12

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

Table B12: Following (by branch). Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5b* in main body)

	Medical science	Humanities	Formal science	Natural science	Social science	No category
Post-COVID	0.716^{***}	0.368^{***}	0.447^{***}	0.286^{***}	0.186^{***}	0.320^{**}
	[0.608, 0.824]	[0.161, 0.575]	[0.253, 0.640]	[0.153, 0.418]	[0.121, 0.251]	[0.114, 0.526]
Num.Obs.	334 670	334 670	334 670	334 670	334 670	334 670
R2 Marg.	0.031	0.009	0.014	0.006	0.002	0.007
R2 Cond.	0.195	0.119	0.081	0.089	0.101	0.125
ICC	0.2	0.1	0.1	0.1	0.1	0.1
RMSE	0.07	0.03	0.04	0.05	0.10	0.03

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

Table B13: Mentioning (by branch). Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5b* in main body)

	Medical science	Humanities	Formal science	Natural science	Social science	No category
Post-COVID	1.124^{***}	0.135	-0.327	0.355^{+}	0.263^{***}	-0.334^{+}
	[0.825, 1.422]	[−0.361, 0.630]	[−0.784, 0.129]	[−0.022, 0.732]	[0.131, 0.396]	[−0.680, 0.012]
Num.Obs.	315 542	315 542	315 542	315 542	315 542	315 542
R2 Marg.	0.077	0.001	0.006	0.008	0.005	0.006
R2 Cond.	0.194	0.366	0.273	0.208	0.104	0.271
ICC	0.1	0.4	0.3	0.2	0.1	0.3
RMSE	0.03	0.01	0.01	0.02	0.05	0.02

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

Table B14: Retweeting (by branch). Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5b* in main body)

	Medical science	Humanities	Formal science	Natural science	Social science	No category
Post-COVID	0.842^{***}	-0.068	-0.288^{***}	0.251^{***}	0.038^{+}	0.030
	[0.741, 0.942]	[−0.237, 0.101]	[−0.452, −0.123]	[0.130, 0.372]	[−0.007, 0.084]	[−0.150, 0.211]
Num.Obs.	925 499	925 499	925 499	925 499	925 499	925 499
R2 Marg.	0.041	0.000	0.005	0.004	0.000	0.000
R2 Cond.	0.219	0.198	0.253	0.237	0.141	0.175
ICC	0.2	0.2	0.2	0.2	0.1	0.2
RMSE	0.05	0.02	0.02	0.03	0.09	0.02

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

Table B15: Quote tweeting (by branch). Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5b* in main body)

	Medical science	Humanities	Formal science	Natural science	Social science	No category
Post-COVID	0.790^{***}	-0.021	-0.061	0.367^{*}	0.012	0.498^{*}
	[0.552, 1.027]	[−0.427, 0.386]	[−0.424, 0.302]	[0.075, 0.658]	[−0.101, 0.126]	[0.048, 0.948]
Num.Obs.	166 371	166 371	166 371	166 371	166 371	166 371
R2 Marg.	0.039	0.000	0.000	0.008	0.000	0.016
R2 Cond.	0.163	0.268	0.273	0.183	0.067	0.137
ICC	0.1	0.3	0.3	0.2	0.1	0.1
RMSE	0.05	0.02	0.03	0.03	0.09	0.02

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

Table B16: Replying tweeting (by branch). Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth (*Fig. 5b* in main body)

	Medical science	Humanities	Formal science	Natural science	Social science	No category
Post-COVID	0.187^{+}	-0.249^{*}	-0.304^{**}	-0.227^{*}	-0.028	0.309^{+}
	[−0.003, 0.377]	[−0.494, −0.003]	[−0.534, −0.074]	[−0.406, −0.049]	[−0.110, 0.053]	[−0.004, 0.621]
Num.Obs.	266 846	266 846	266 846	266 846	266 846	266 846
R2 Marg.	0.002	0.003	0.005	0.003	0.000	0.005
R2 Cond.	0.309	0.248	0.265	0.311	0.271	0.304
ICC	0.3	0.2	0.3	0.3	0.3	0.3
RMSE	0.04	0.03	0.03	0.04	0.09	0.02

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

Appendix C Supporting analyses

C.1 Estimated effects of legislator and legislature characteristics on the proportion of researchers on behaviors

These are results from linear mixed-effects models with country of legislature random effects with age (under 35), party family (other), country region (Europe), system (presidential), and Q1 for followers and friends as references for categorical variables. Panel a presents the coefficients with 80% and 95% confidence intervals. The conditional predictions are computed with numeric covariates are held at their means and the other covariates at their modes: no research degree, presidential, European, male, 45-54, Conservative party, and the first quartile of followers and friends.

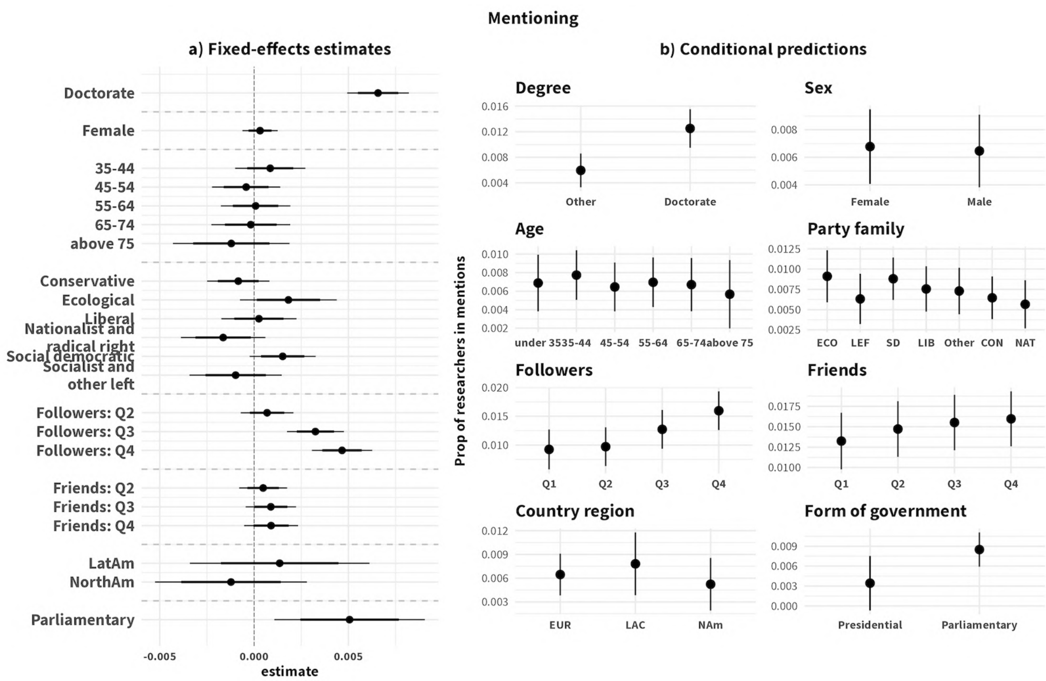


Figure C1: Estimated effects of legislator and legislature characteristics on the proportion of mentions of researchers

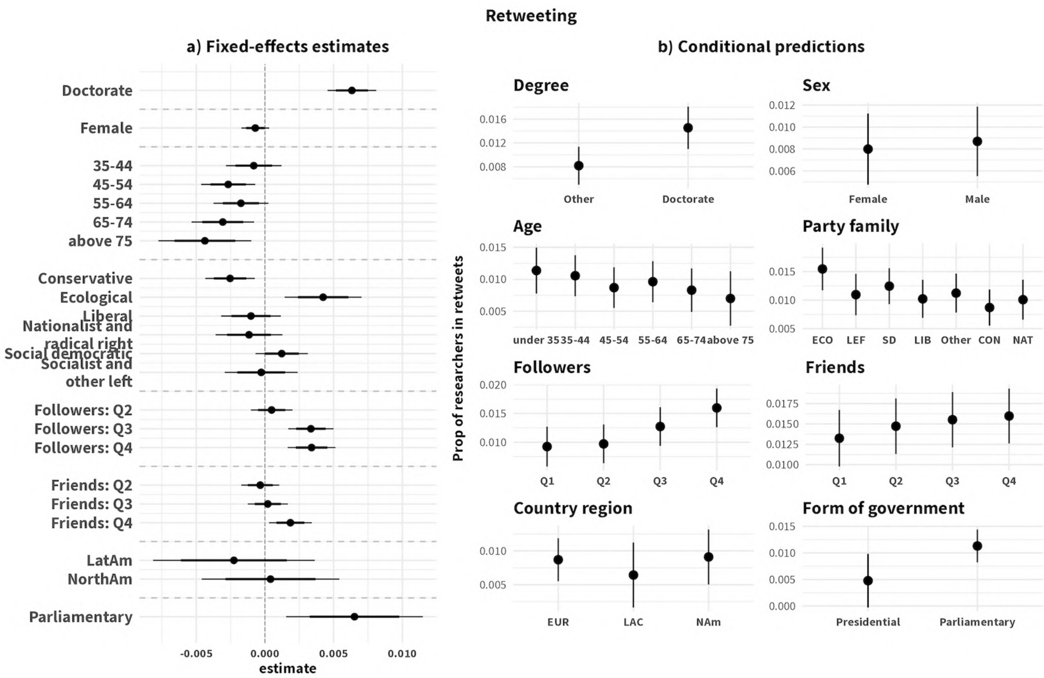


Figure C2: Estimated effects of legislator and legislature characteristics on the proportion of retweeting of researchers

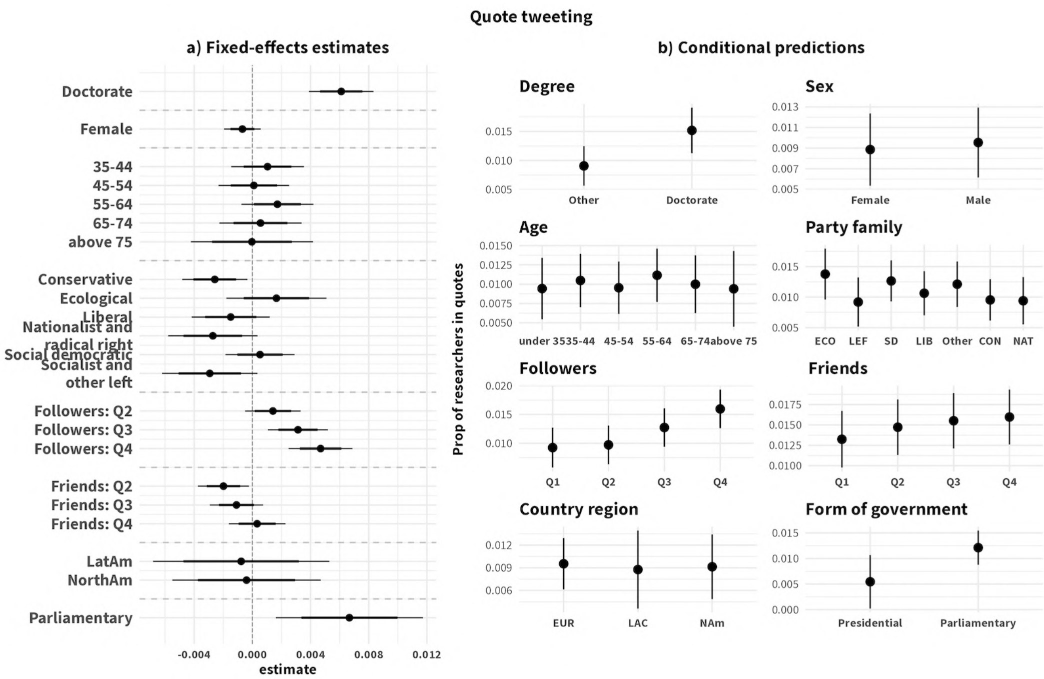


Figure C3: Estimated effects of legislator and legislature characteristics on the proportion of quoting of researchers

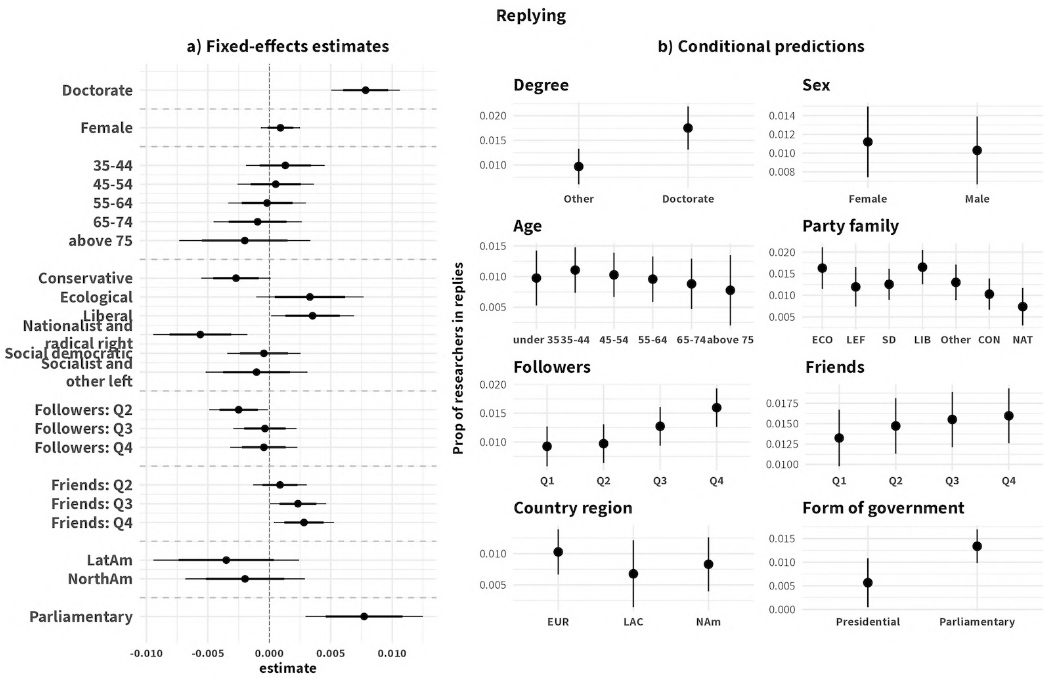


Figure C4: Estimated effects of legislator and legislature characteristics on the proportion of replies to researchers

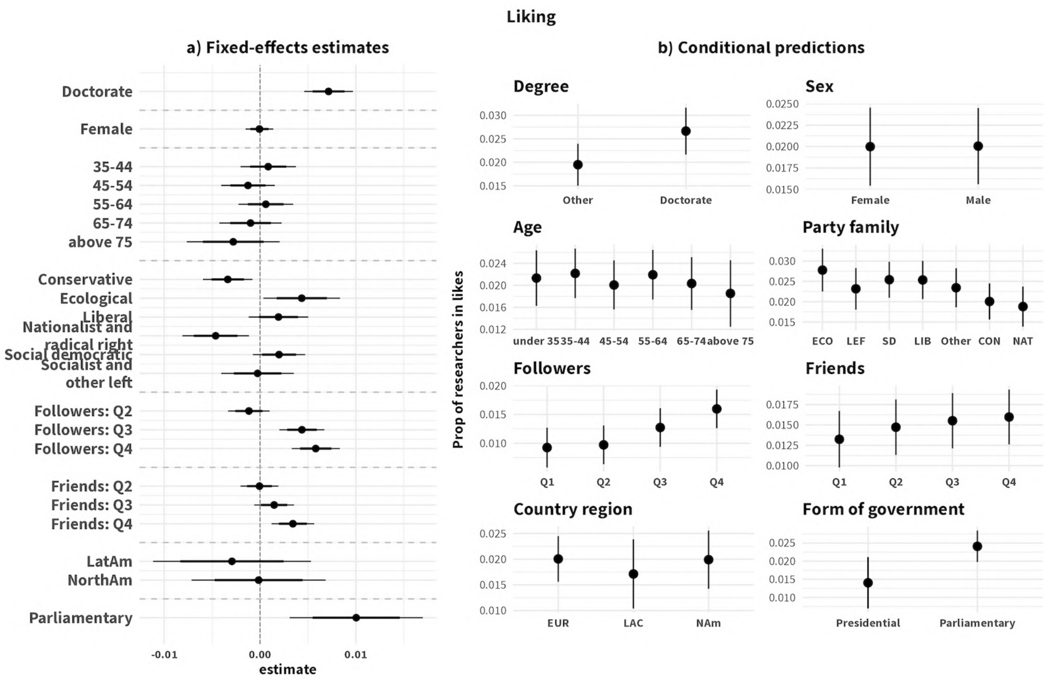


Figure C5: Estimated effects of legislator and legislature characteristics on the proportion of likes of researcher Tweets

C.2 Estimated effects of legislator and legislature characteristics on the absolute number of researchers on behaviors

These are results from linear mixed-effects models with legislature random effects with age (under 35), party family (other), country region (Europe), system (presidential), and Q1 for followers and friends as references for categorical variables. Panel a presents the coefficients with 80% and 95% confidence intervals. The conditional predictions are computed with numeric covariates are held at their means and the other covariates at their modes: no research degree, presidential, European, male, 45-54, Conservative party, and the first quartile of followers and friends.

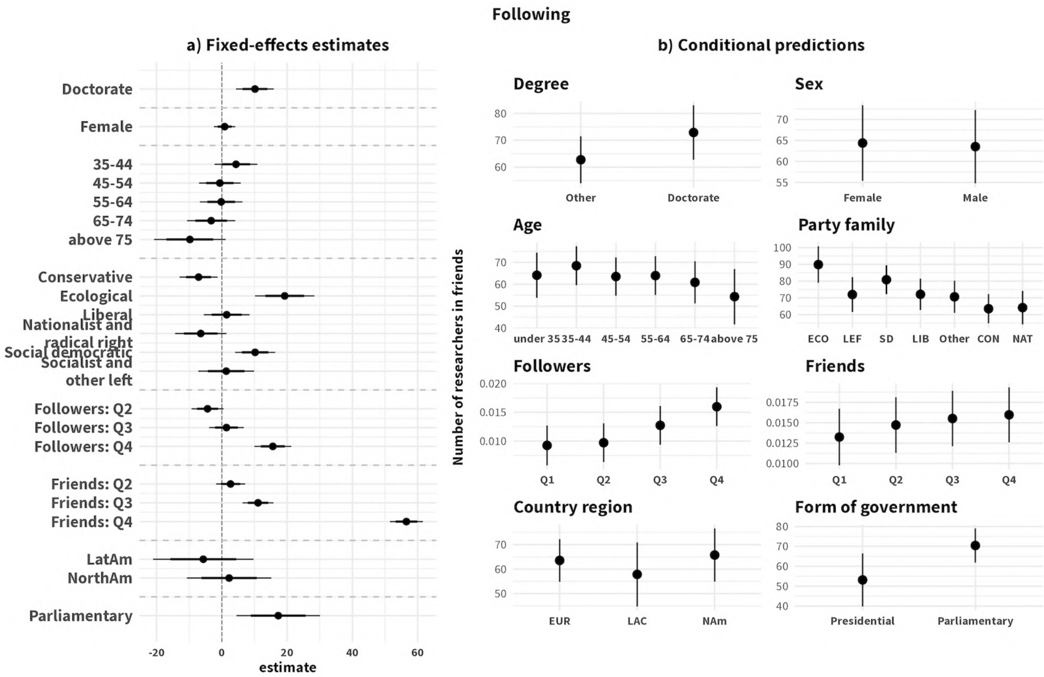


Figure C6: Estimated effects of legislator and legislature characteristics on the absolute number of friends of researcher Tweets

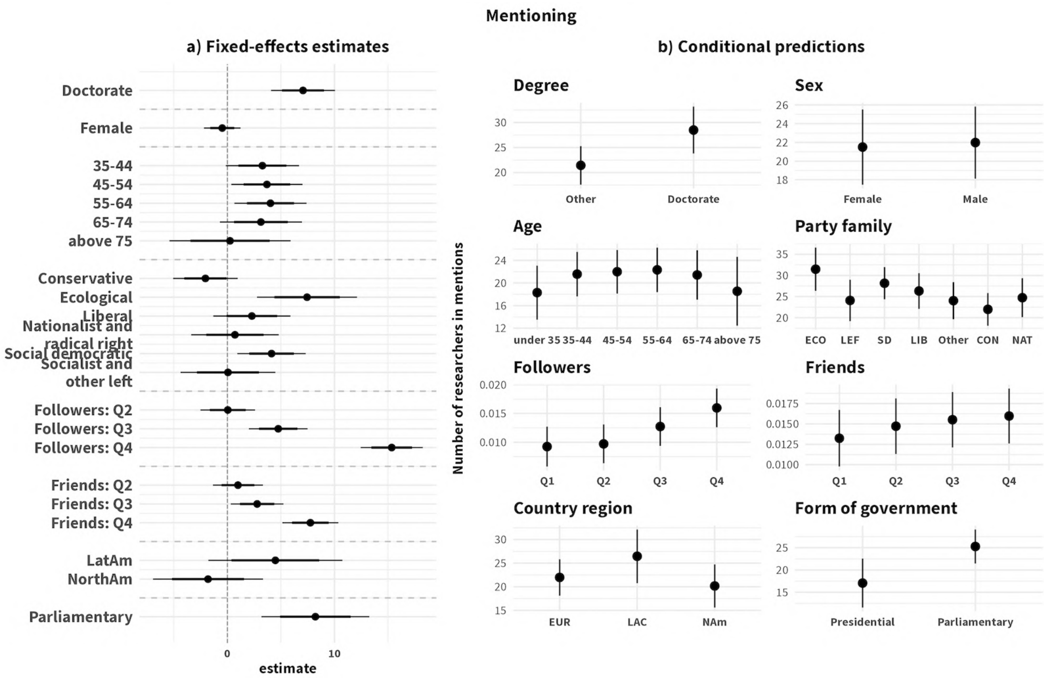


Figure C7: **Estimated effects of legislator and legislature characteristics on the absolute number of mentions of researchers**

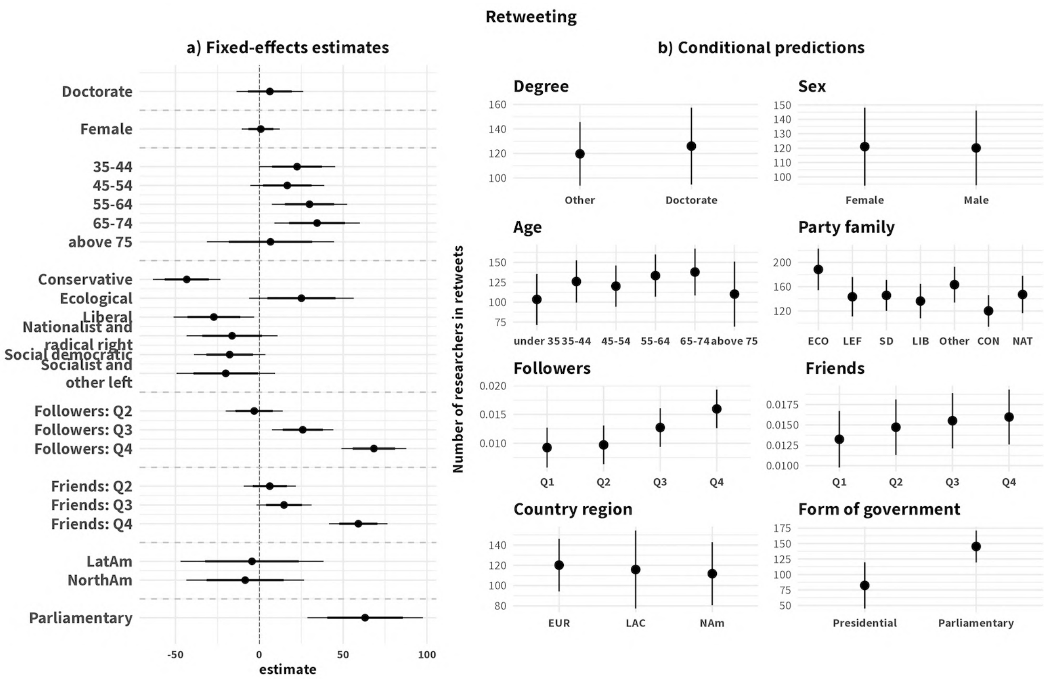


Figure C8: **Estimated effects of legislator and legislature characteristics on the absolute number of retweeting of researchers**

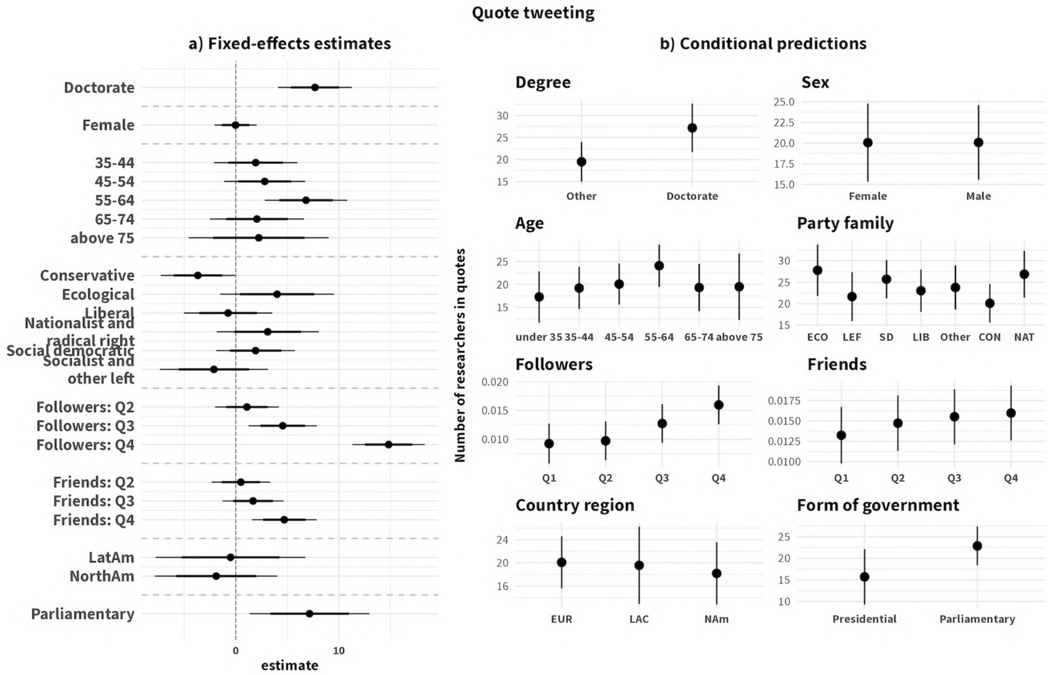


Figure C9: Estimated effects of legislator and legislature characteristics on the absolute number of quoting of researchers

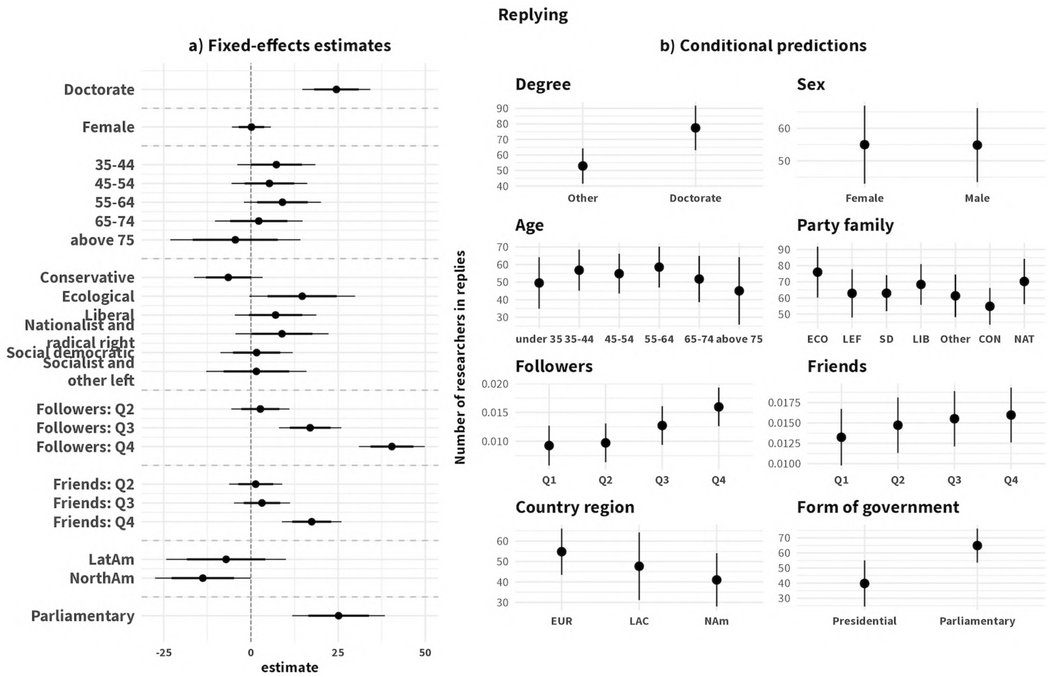


Figure C10: Estimated effects of legislator and legislature characteristics on the absolute number of replies to researchers

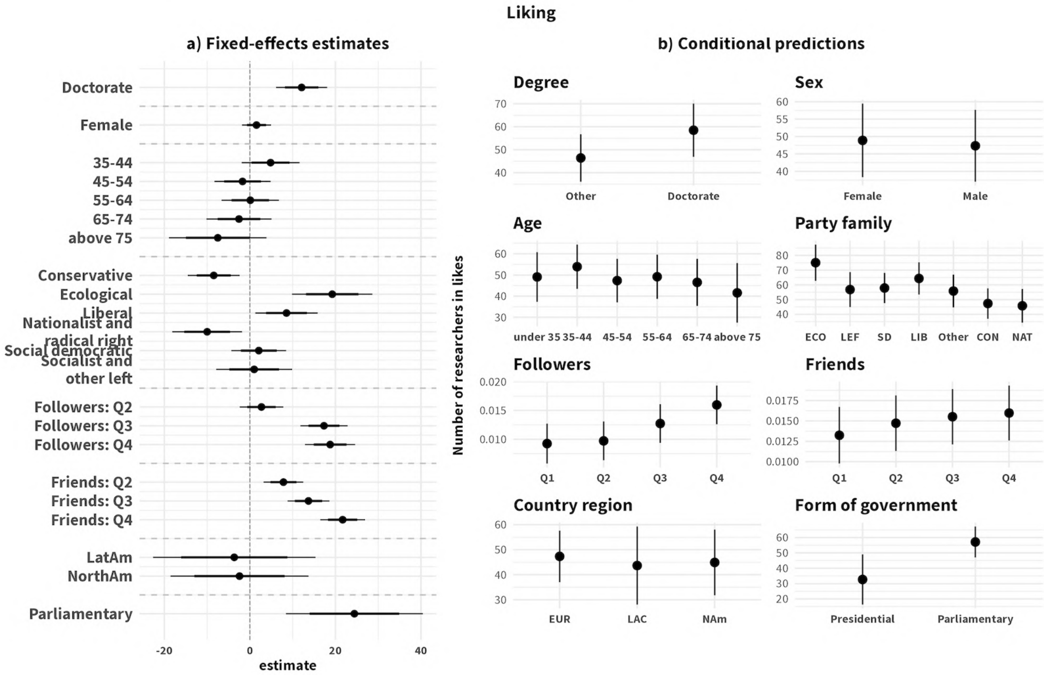


Figure C11: Estimated effects of legislator and legislature characteristics on the absolute number of likes of researcher Tweets

Table C1: Estimated effects of legislator and legislature characteristics on the absolute number of researchers across behaviors. (*Figs. C6-C11*)

	Following	Mentioning	Retweeting	Quote tweeting	Replying	Liking
Doctoral studies	10.073***	7.076***	6.422	7.708***	24.598***	12.123***
	[4.286, 15.861]	[4.086, 10.065]	[−13.515, 26.359]	[4.121, 11.294]	[14.828, 34.368]	[6.156, 18.090]
Female	0.825	−0.489	0.902	−0.023	0.164	1.561
	[−2.457, 4.107]	[−2.196, 1.219]	[−10.454, 12.258]	[−2.062, 2.016]	[−5.415, 5.743]	[−1.828, 4.951]
Age: 35-44 (ref: Under 35)	4.271	3.278^{+}	22.643^{+}	1.943	7.352	4.869
	[−2.312, 10.853]	[−0.144, 6.700]	[−0.121, 45.406]	[−2.111, 5.997]	[−3.848, 18.552]	[−1.909, 11.647]
45-54	−0.592	3.705*	16.737	2.817	5.316	−1.708
	[−6.988, 5.804]	[0.381, 7.029]	[−5.389, 38.862]	[−1.118, 6.752]	[−5.574, 16.205]	[−8.285, 4.868]
55-64	−0.183	4.044*	30.025**	6.837***	9.073	0.112
	[−6.679, 6.313]	[0.668, 7.421]	[7.546, 52.504]	[2.837, 10.836]	[−1.987, 20.133]	[−6.574, 6.799]
65-74	−3.006	3.175	34.411**	2.074	2.162	−2.589
	[−10.382, 4.371]	[−0.654, 7.004]	[8.929, 59.893]	[−2.494, 6.642]	[−10.397, 14.721]	[−10.195, 5.017]
above 75	−9.185	0.334	6.518	2.275	−4.675	−7.619
	[−20.149, 1.779]	[−5.321, 5.989]	[−31.346, 44.382]	[−4.508, 9.059]	[−23.316, 13.966]	[−18.987, 3.748]
Party family: Conservative (ref: Other)	−7.146*	−2.035	−43.183***	−3.705*	−6.479	−8.457**
	[−12.976, −1.316]	[−5.051, 0.982]	[−63.299, −23.068]	[−7.319, −0.092]	[−16.326, 3.368]	[−14.523, −2.392]
Ecological	19.246***	7.459**	25.138	3.983	14.732^{+}	19.198***
	[10.148, 28.344]	[2.784, 12.133]	[−6.054, 56.331]	[−1.549, 9.515]	[−0.490, 29.954]	[9.793, 28.603]
Liberal	1.507	2.209	−27.616*	−0.845	6.923	8.524*
	[−5.527, 8.541]	[−1.411, 5.829]	[−51.719, −3.512]	[−5.156, 3.467]	[−4.817, 18.663]	[1.208, 15.839]
Nationalistic and radical right	−6.485	0.696	−16.318	3.071	8.947	−9.978*
	[−14.359, 1.389]	[−3.395, 4.788]	[−43.489, 10.852]	[−1.883, 8.024]	[−4.426, 22.321]	[−18.145, −1.812]
Social democratic	10.046**	4.158*	−17.395	1.921	1.781	2.107
	[3.880, 16.212]	[0.969, 7.347]	[−38.687, 3.897]	[−1.901, 5.744]	[−8.619, 12.180]	[−4.305, 8.519]
Socialist and other left	1.292	0.013	−20.264	−2.198	1.525	1.012
	[−7.261, 9.846]	[−4.424, 4.450]	[−49.663, 9.134]	[−7.455, 3.058]	[−12.875, 15.924]	[−7.880, 9.903]
Followers qtl: Q2 (ref: Q1)	−4.372^{+}	0.033	−3.102	1.070	2.651	2.696
	[−9.281, 0.536]	[−2.519, 2.585]	[−20.100, 13.896]	[−2.018, 4.158]	[−5.700, 11.002]	[−2.406, 7.798]
Q3	1.270	4.762***	26.101**	4.551**	17.028***	17.354***
	[−4.036, 6.576]	[2.015, 7.510]	[7.795, 44.407]	[1.234, 7.868]	[8.075, 25.981]	[11.829, 22.880]
Q4	16.008***	15.416***	68.273***	14.861***	40.305***	18.705***
	[10.383, 21.633]	[12.510, 18.323]	[48.904, 87.642]	[11.346, 18.377]	[30.849, 49.762]	[12.848, 24.563]
Friends qtl: Q2 (ref: Q1)	2.734	0.990	6.189	0.495	1.335	7.836***
	[−1.737, 7.206]	[−1.342, 3.321]	[−9.335, 21.713]	[−2.345, 3.334]	[−6.310, 8.980]	[3.187, 12.484]
Q3	11.063***	2.765*	14.677^{+}	1.664	3.157	13.704***
	[6.334, 15.792]	[0.310, 5.220]	[−1.697, 31.051]	[−1.301, 4.630]	[−4.884, 11.198]	[8.808, 18.599]
Q4	56.514***	7.732***	58.949***	4.691**	17.403***	21.699***
	[51.470, 61.557]	[5.121, 10.343]	[41.525, 76.373]	[1.542, 7.839]	[8.862, 25.944]	[16.471, 26.926]
Region: LAC (ref: EUR)	−5.018	4.128	−7.721	−0.800	−7.743	−3.969
	[−21.107, 11.070]	[−2.531, 10.788]	[−51.970, 36.527]	[−8.546, 6.946]	[−25.776, 10.289]	[−24.545, 16.607]
NAm	1.209	−1.326	−3.874	−1.493	−12.821^{+}	−1.972
	[−12.668, 15.086]	[−6.949, 4.297]	[−41.194, 33.447]	[−7.999, 5.013]	[−27.591, 1.948]	[−19.943, 16.000]
Parliamentary republic	18.249**	7.784**	58.836**	6.787*	24.293***	23.998**
	[4.546, 31.952]	[2.273, 13.295]	[22.273, 95.399]	[0.410, 13.164]	[9.974, 38.612]	[6.166, 41.830]
SD (Observations)	44.711	23.003	153.695	26.939	75.272	45.639
Num.Obs.	3381	3318	3335	3145	3311	3295
R2 Marg.	0.310	0.139	0.107	0.086	0.095	0.163
R2 Cond.	0.330	0.154	0.122	0.101	0.103	0.203
ICC	0.0	0.0	0.0	0.0	0.0	0.0
RMSE	44.52	22.90	153.04	26.82	74.96	45.43

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.3 Auxiliary analyses for pre- and during COVID legislator-researcher engagement comparison

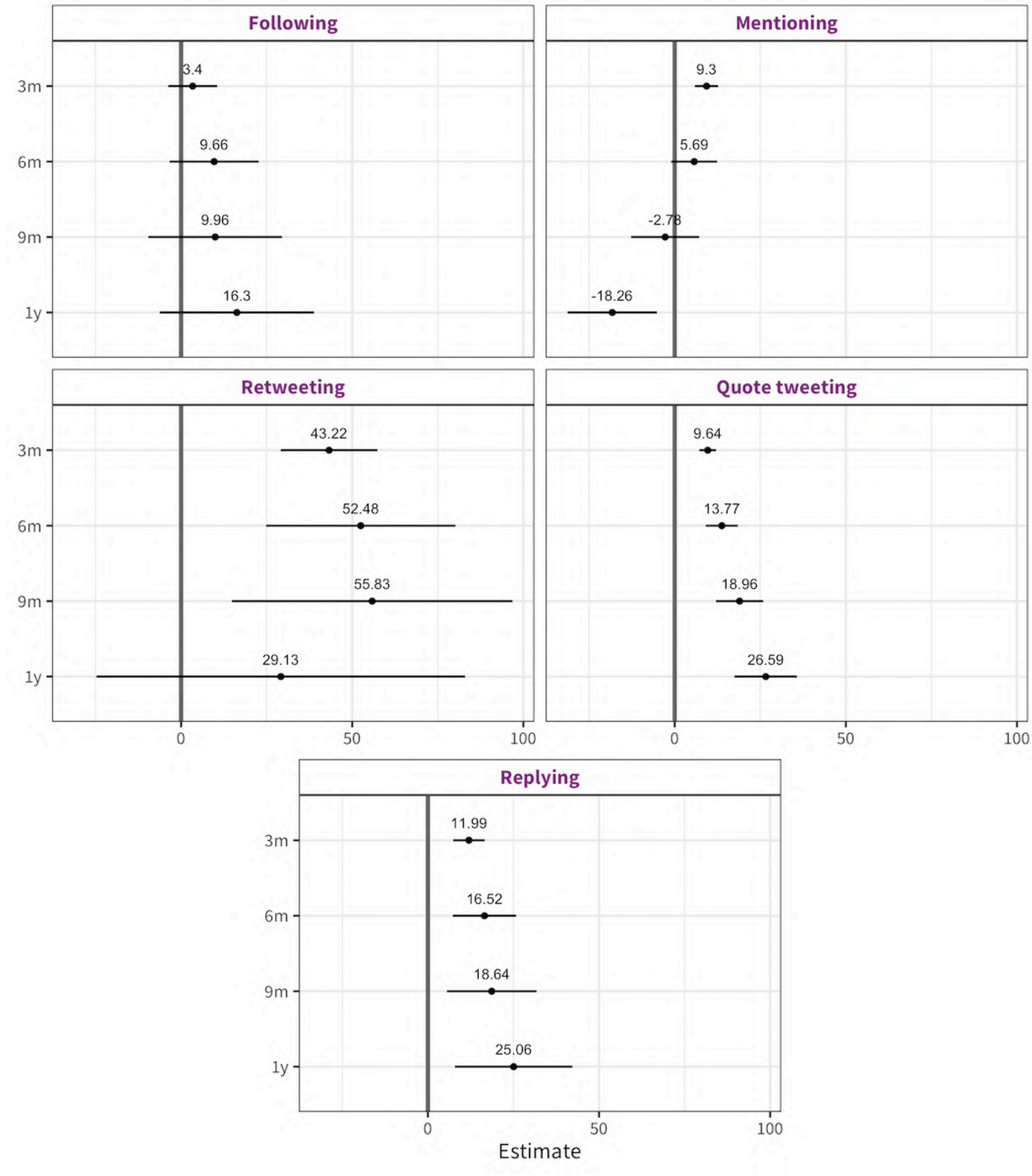


Figure C12: Differences-in-means between pre- and during-COVID number of behaviors. Results from OLS estimates

Table C2: Differences-in-means between pre- and during-COVID number of behaviors across bandwidths. Results from OLS estimates (*Fig. C12*)

Behavior	Estimate	Std. error	P. value	Conf. int
Following (3m)	3.48	3.66	0.34	[-3.7,10.7]
Following (6m)	9.79	6.65	0.14	[-3.3,22.8]
Following (9m)	10.08	9.96	0.31	[-9.5,29.6]
Following (1y)	16.44	11.53	0.15	[-6.2,39]
Mentioning (3m)	10.23	1.75	0.00	[6.8,13.7]
Mentioning (6m)	7.37	3.47	0.03	[0.6,14.2]
Mentioning (9m)	-0.39	5.12	0.94	[-10.4,9.6]
Mentioning (1y)	-15.25	6.73	0.02	[-28.4,-2.1]
Retweeting (3m)	46.42	7.29	0.00	[32.1,60.7]
Retweeting (6m)	58.35	14.28	0.00	[30.4,86.3]
Retweeting (9m)	64.08	21.15	0.00	[22.6,105.5]
Retweeting (1y)	39.47	27.78	0.16	[-15,93.9]
Quote tweeting (3m)	10.12	1.25	0.00	[7.7,12.6]
Quote tweeting (6m)	14.67	2.42	0.00	[9.9,19.4]
Quote tweeting (9m)	20.28	3.56	0.00	[13.3,27.3]
Quote tweeting (1y)	28.39	4.69	0.00	[19.2,37.6]
Replying (3m)	12.74	2.38	0.00	[8.1,17.4]
Replying (6m)	17.90	4.77	0.00	[8.5,27.3]
Replying (9m)	20.59	6.75	0.00	[7.4,33.8]
Replying (1y)	27.66	8.85	0.00	[10.3,45]

Table C3: Results from logistic regression on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth with party family interaction (*Fig. C13*)

	Following	Mentioning	Retweeting	Quote tweeting	Replying
Post-COVID	0.391^{***}	0.362^{**}	0.399^{***}	0.185^{+}	-0.079
	[0.287, 0.495]	[0.113, 0.615]	[0.308, 0.491]	[−0.019, 0.393]	[−0.225, 0.067]
Ecological	1.026^{***}	1.420^{***}	1.478^{***}	0.960^{***}	1.037^{***}
	[0.892, 1.159]	[1.077, 1.750]	[1.359, 1.597]	[0.656, 1.252]	[0.871, 1.201]
Liberal	0.194^{**}	0.663^{***}	0.556^{***}	0.017	0.305^{***}
	[0.066, 0.320]	[0.369, 0.954]	[0.429, 0.681]	[−0.300, 0.319]	[0.131, 0.477]
Nationalist and radical right	-0.973^{***}	-1.017^{*}	-0.391^{***}	0.260	-0.932^{***}
	[−1.300, −0.674]	[−2.201, −0.148]	[−0.587, −0.205]	[−0.077, 0.579]	[−1.320, −0.581]
Other	0.274^{***}	0.749^{***}	0.975^{***}	0.832^{***}	0.146
	[0.132, 0.412]	[0.437, 1.055]	[0.877, 1.073]	[0.589, 1.073]	[−0.040, 0.328]
Social democratic	0.618^{***}	1.042^{***}	1.101^{***}	0.577^{***}	0.419^{***}
	[0.525, 0.712]	[0.814, 1.277]	[1.014, 1.188]	[0.385, 0.774]	[0.291, 0.549]
Socialist and other left	-0.034	0.087	0.049	-0.255	0.004
	[−0.207, 0.132]	[−0.364, 0.502]	[−0.139, 0.230]	[−0.742, 0.180]	[−0.254, 0.248]
Post-COVID*Ecological	-0.108	0.064	-0.093	0.066	-0.043
	[−0.289, 0.073]	[−0.365, 0.497]	[−0.241, 0.055]	[−0.306, 0.443]	[−0.269, 0.183]
Post-COVID*Liberal	0.141^{+}	-0.078	-0.307^{***}	0.219	0.232^{*}
	[−0.026, 0.308]	[−0.451, 0.296]	[−0.462, −0.150]	[−0.149, 0.596]	[0.009, 0.456]
Post-COVID*Nationalist and radical right	-0.060	-0.415	-0.428^{***}	-0.522^{*}	-0.054
	[−0.460, 0.350]	[−1.877, 1.048]	[−0.676, −0.177]	[−0.973, −0.072]	[−0.560, 0.456]
Post-COVID*Other	0.121	0.070	-0.130^{*}	-0.227	0.130
	[−0.060, 0.302]	[−0.329, 0.471]	[−0.253, −0.006]	[−0.551, 0.096]	[−0.119, 0.380]
Post-COVID*Social democratic	-0.219^{***}	-0.164	-0.383^{***}	0.070	-0.100
	[−0.347, −0.092]	[−0.463, 0.133]	[−0.492, −0.274]	[−0.178, 0.316]	[−0.280, 0.079]
Post-COVID*Socialist and other left	-0.212^{+}	-0.199	-0.092	0.069	0.004
	[−0.455, 0.031]	[−0.770, 0.380]	[−0.322, 0.141]	[−0.497, 0.659]	[−0.326, 0.338]
Num.Obs.	334 670	315 542	925 499	166 371	266 846
RMSE	0.15	0.07	0.11	0.11	0.12

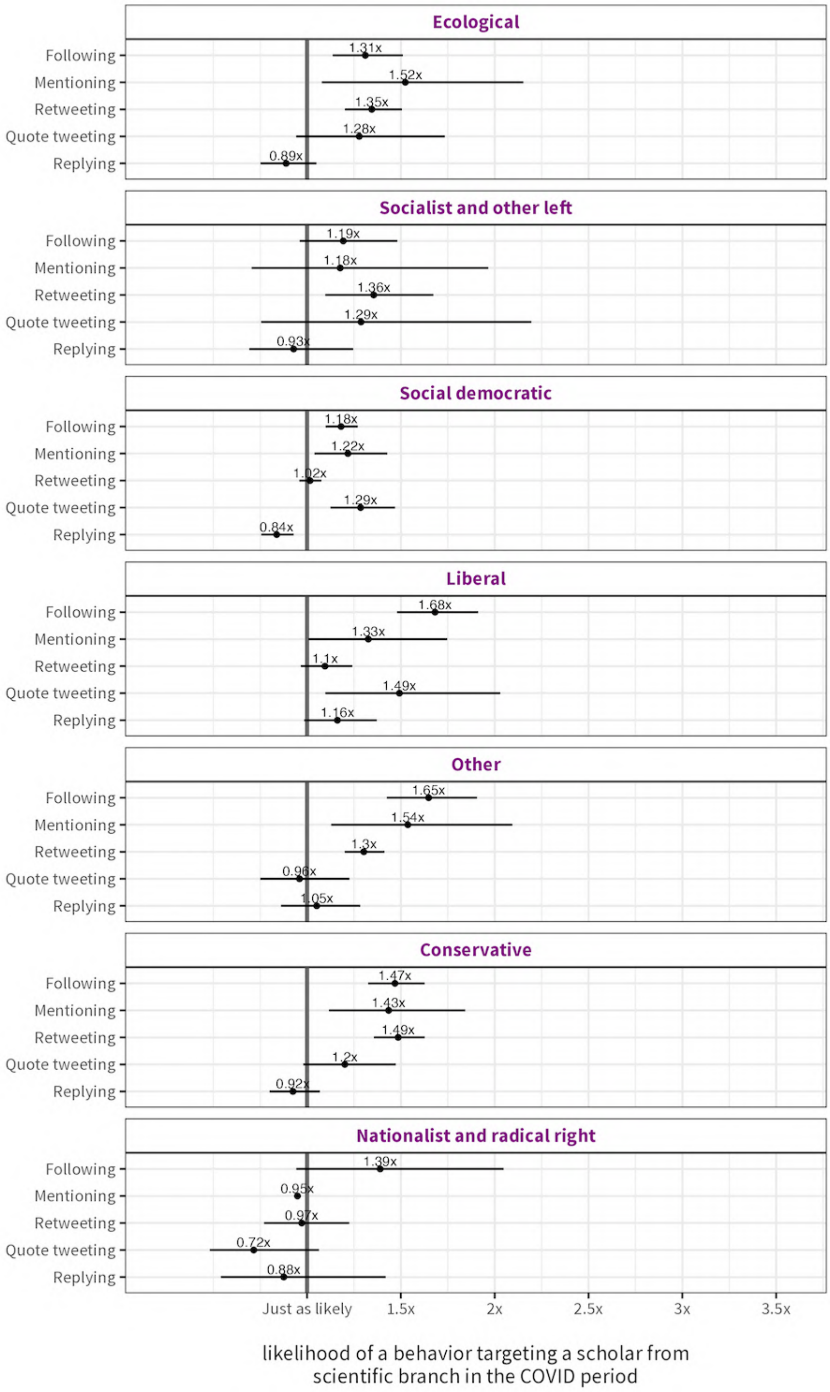
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C4: Results from a logistic mixed effects model with country of legislature random effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a 12 week bandwidth with Doctorate interaction (*Fig. C14*)

	Following	Mentioning	Retweeting	Quote tweeting	Replying
Post-COVID	0.313^{***}	0.351^{***}	0.177^{***}	0.190^{***}	-0.032
	[0.262, 0.363]	[0.242, 0.461]	[0.141, 0.214]	[0.094, 0.285]	[−0.101, 0.036]
Doctorate	0.412^{***}	0.667^{***}	0.344^{***}	0.534^{***}	0.710^{***}
	[0.296, 0.529]	[0.456, 0.878]	[0.276, 0.412]	[0.324, 0.745]	[0.584, 0.836]
Post-COVID*Doctorate	0.201^{**}	-0.335^{*}	-0.175^{***}	-0.050	-0.109
	[0.053, 0.348]	[−0.606, −0.063]	[−0.259, −0.090]	[−0.308, 0.208]	[−0.272, 0.055]
Num.Obs.	313 252	315 542	925 499	166 371	266 846
R2 Marg.	0.012	0.013	0.003	0.008	0.008
R2 Cond.	0.097	0.062	0.146	0.076	0.264
ICC	0.1	0.0	0.1	0.1	0.3
RMSE	0.15	0.07	0.11	0.11	0.12

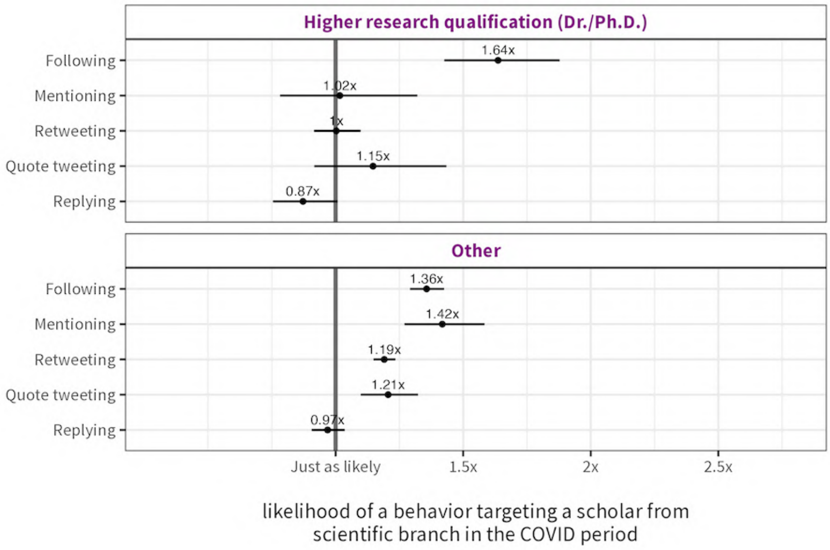
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure C13: Subgroup marginal effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a ±12 week bandwidth (Party family).



Note. Results from logistic regression model. The estimates in the figure are relative risks representing the ratio of the probability of an event in the COVID period to the probability of an outcome in a pre-COVID period.

Figure C14: Subgroup marginal effects on following and engagement with academic researchers during the COVID versus pre-COVID periods with a ±12 week bandwidth (Research qualifications).



Note. Results from a logistic mixed-effects models with country of legislature random effects. The estimates in the figure are relative risks representing the ratio of the probability of an event in the COVID period to the probability of an outcome in a pre-COVID period.

Appendix D Software statement

I used R version 4.2.1 (R Core Team, 2022) and the following R packages:

`academictwitteR` v. 0.3.1 (Barrie and Ho, 2021)
`data.table` v. 1.14.2 (Dowle and Srinivasan, 2021)
`estimatr` v. 1.0.0 (Blair et al., 2022)
`ggh4x` v. 0.2.6 (van den Brand, 2023)
`gt` v. 0.9.0 (Iannone et al., 2023)
`gtExtras` v. 0.5.0.9000 (Mock, 2023)
`janitor` v. 2.1.0 (Firke, 2021)
`kableExtra` v. 1.3.4 (Zhu, 2021)
`lme4` v. 1.1.31 (Bates, Mächler, et al., 2015)
`marginaleffects` v. 0.11.1 (Arel-Bundock, 2023)
`Matrix` v. 1.5.3 (Bates, Maechler, and Jagan, 2022)
`modelsummary` v. 1.4.3 (Arel-Bundock, 2022)
`pacman` v. 0.5.1 (Rinker and Kurkiewicz, 2018)
`patchwork` v. 1.1.2 (Pedersen, 2022)
`psych` v. 2.2.5 (Revelle, 2022)
`reactable` v. 0.4.4 (Lin, 2023)
`reactablefmtr` v. 2.0.0 (Cuilla, 2022)
`rtweet` v. 1.0.2 (Kearney, 2019b)
`scales` v. 1.3.0 (Wickham, Pedersen, and Seidel, 2023)
`tidyverse` v. 2.0.0 (Wickham, Averick, et al., 2019)
`urltools` v. 1.7.3 (Keyes et al., 2019)
`webshot2` v. 0.1.0 (Chang, 2022)
`WikidataR` v. 2.3.3 (Shafee, Keyes, and Signorelli, 2021)
`xtable` v. 1.8.4 (Dahl et al., 2019)

References

Arel-Bundock, Vincent (2022). “modelsummary: Data and Model Summaries in R”. *Journal of Statistical Software* 103.1, pp. 1–23.

— (2023). *marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*.

Barrie, Christopher and Justin Chun-ting Ho (2021). “academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint”. *Journal of Open Source Software* 6.62, p. 3272.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. *Journal of Statistical Software* 67.1, pp. 1–48.

Bates, Douglas, Martin Maechler, and Mikael Jagan (2022). *Matrix: Sparse and Dense Matrix Classes and Methods*.

Blair, Graeme, Jasper Cooper, Alexander Coppock, Macartan Humphreys, and Luke Sonnet (2022). *estimatr: Fast Estimators for Design-Based Inference*.

Chang, Winston (2022). *webshot2: Take Screenshots of Web Pages*.

Cuilla, Kyle (2022). *reactablefmtr: Streamlined Table Styling and Formatting for Reactable*.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton (2019). *xtable: Export Tables to LaTeX or HTML*.

Dowle, Matt and Arun Srinivasan (2021). *data.table: Extension of ‘data.frame’*.

Firke, Sam (2021). *janitor: Simple Tools for Examining and Cleaning Dirty Data*.

Göbel, Sascha and Simon Munzert (2022). “The Comparative Legislators Database”. *British Journal of Political Science* 52.3, pp. 1398–1408.

“International Standard Classification of Education, ISCED 1997” (2003). In: *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*. Ed. by Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf. Boston, MA: Springer US, pp. 195–220.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, and JooYoung Seo (2023). *gt: Easily Create Presentation-Ready Display Tables*.

Kearney, Michael W. (2019a). “rtweet: Collecting and analyzing Twitter data”. *Journal of Open Source Software* 4.42, p. 1829.

— (2019b). “rtweet: Collecting and analyzing Twitter data”. *Journal of Open Source Software* 4.42, p. 1829.

Keyes, Os, Jay Jacobs, Drew Schmidt, Mark Greenaway, Bob Rudis, Alex Pinto, Maryam Khezrzadeh, Peter Meilstrup, Adam M. Costello, Jeff Bezanson, Peter Meilstrup, and Xueyuan Jiang (2019). *urltools: Vectorised Tools for URL Handling and Parsing*.

Lehmann, Pola, Simon Franzmann, Tobias Burst, Sven Regel, Felicia Riethmüller, Andrea Volkens, Bernhard Weßels, and Lisa Zehnter (2023). *The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2023a*.

Lin, Greg (2023). *reactable: Interactive Data Tables for R*.

Meeder, Brendan, Brian Karrer, Amin Sayedi, R. Ravi, Christian Borgs, and Jennifer Chayes (2011). “We know who you followed last summer: inferring social link creation times in twitter”. In: *Proceedings of the 20th international conference on World wide web*. WWW ’11. New York, NY, USA: Association for Computing Machinery, pp. 517–526.

Mock, Thomas (2023). *gtExtras: Extending ‘gt’ for Beautiful HTML Tables*.

Mongeon, Philippe, Timothy D. Bowman, and Rodrigo Costas (2023). “An open data set of scholars on Twitter”. *Quantitative Science Studies*, pp. 1–11.

Pedersen, Thomas Lin (2022). *patchwork: The Composer of Plots*.

Priem, Jason, Heather Piwowar, and Richard Orr (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Revelle, William (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. Evanston, Illinois.

Rinker, Tyler W. and Dason Kurkiewicz (2018). *pacman: Package Management for R*. Buffalo, New York.

Shafee, Thomas, Os Keyes, and Serena Signorelli (2021). *WikidataR: Read-Write API Client Library for Wikidata*.

Shen, Zhihong, Hao Ma, and Kuansan Wang (2018). *A Web-scale system for scientific knowledge exploration*.

van den Brand, Teun (2023). *ggh4x: Hacks for 'ggplot2'*.
Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani (2019). "Welcome to the tidyverse". *Journal of Open Source Software* 4.43, p. 1686.
Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel (2023). *scales: Scale Functions for Visualization*.
Wojcieszak, Magdalena, Andreu Casas, Xudong Yu, Jonathan Nagler, and Joshua A. Tucker (2022). "Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity". *Science Advances* 8.39, eabn9418.
Zhu, Hao (2021). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*.

Policy documents across 185 countries predominantly rely on evidence from the Global North

Sebastian Ramirez-Ruiz (Hertie School)*

Roman Senninger (Aarhus University)

Abstract. Evidence is widely acknowledged as essential for crafting effective public policies. Despite its critical role, we know surprisingly little about the specific sources that inform decisions around the world. This paper explores the sources of evidence in the policymaking arena by analyzing evidence cited in over 1.2 million policy documents from 185 countries. Our analyses capture references to 3.5 million scholarly works and 740,000 policy sources including contributions from government agencies, academic researchers, international organizations, and think tanks. By focusing on the *documented*, *accessible*, and *digitally visible* evidence available to policymakers, we map global patterns of evidence use, highlighting regional and policy domain variation. Our findings reveal a pronounced concentration of attention: the vast majority of cited evidence—both academic and policy—is produced in the Global North, even in documents authored by governments in the Global South. These patterns persist across policy areas, though with notable variation in the types of sources commonly used. Overall, the findings reveal a highly concentrated evidence landscape, where a small number of countries disproportionately serve as global reference points, underscoring persistent asymmetries in visibility, influence, and access within the international policy knowledge ecosystem.

Keywords. Evidence-informed policymaking; Expert-based information; Information diffusion; Science-policy nexus; Global policy analysis; Knowledge brokers

*Corresponding author. Address: Friedrichstrasse 180, 10117 Berlin, Germany. Email: ramirez-ruiz@hertie-school.org, Web: seramirezruiz.github.io.

Over the past decades, the principle of evidence-informed policymaking has become popular across policy circles. Guided by the notion that policymaking should be informed by solid evidence (broadly defined as a knowledge claim that is backed up by a recognized scientific method), many governments have increasingly adopted mechanisms to better integrate research, policy analysis, and evaluation into decision-making processes (Kenny et al., 2017; Commission, 2014; United States Congress, 2019; African Union, 2021). Despite widespread acknowledgment of evidence as a foundation of policymaking, surprisingly little is known about the types and origins of the evidence that makes it into the policy arena (National Research Council, 2012). Although government agencies, research institutions, international organizations, and think tanks around the globe all engage in the creation of new knowledge, the extent to which different actors contribute to policymaking remains largely unexplored. Still, understanding the sources informing policymaking is crucial, as information plays a central role in many influential theories of the policy process (Campbell, 2002; Jones and Baumgartner, 2005; Campbell and Pedersen, 2014; Baumgartner and Jones, 2015).

While the principle of evidence-informed policymaking is widely embraced, its implementation is not value-neutral. Access to, transfer of, and processing of information are not merely technical matters in policymaking—they are deeply tied to power dynamics (Schattschneider, 1960; Bachrach and Baratz, 1962; Jones and Baumgartner, 2005). Ultimately, information can determine which problems are prioritized, how policy options are framed, and whose perspectives are represented (Simon, 1944; Lindblom, 1959; Kingdon, 1984). Expert information, in particular, can be used to mobilize interests, reinforce or challenge dominant views, and serve as a strategic resource for coalition-building and contestation in policymaking (Weible, 2008; Campbell and Pedersen, 2014). This underscores the importance of understanding whose knowledge is seen as legitimate and how certain types of expert information gain reach in policy discussions.

Consequently, a key issue in understanding evidence use on a global scale is that not all sources hold the same influence. Certain actors—whether prominent governments, research institutions, or international organizations—might play a disproportionate role in informing policymaking. This can create asymmetries in whose expertise feeds into policy, potentially reinforcing existing geopolitical hierarchies (Kuhlmann et al., 2020). Despite the growing availability of evidence, potential imbalances in the evidence base used might mean that some voices remain underrepresented, and knowledge circulates not necessarily because it is the most fitting, but because it originates from politically or economically dominant actors.

The global landscape of knowledge production and circulation is already shaped by differences in influence and recognition. Research suggests that academic evidence generated in and about the Global North is often treated as universally applicable, whereas knowledge from the Global South is perceived as context-specific and 'localized' (Castro Torres and Alburez-Gutierrez, 2022). These asymmetries are present in scientific practice, where leading nations in global science receive a disproportionate share of citations, even when producing research of comparable quality (Nielsen and Andersen, 2021; Gomez, Herman, and Parigi, 2022). This raises the question of whether similar disparities also shape the use of evidence in policymaking.

One of the most tangible ways to assess the relationship between evidence and policymaking is through the analysis of policy documents. These texts, ranging from white papers to regulatory impact assessments, are widespread artifacts of modern policy and knowledge dissemination practice across the world. More than just passive repositories of information, policy documents can act as signaling devices, reflecting underlying institutional preferences and evolving norms around what counts as legitimate evidence (Freeman and Maybin, 2011). While there has been valuable research examining the use of evidence and references in policy documents, much of this work remains fragmented. Studies employing policy documents often focus narrowly on the role of scientific research, are confined to specific policy domains like climate change, or predominantly analyze cases from the U.S. (e.g., Bornmann, Haunschild, and Marx, 2016; Isett, Hicks, and Kingsley, 2025; Furnas, LaPira, and Wang, 2025; Ma and Cheng, 2025). A broader, cross-national perspective on how different sources of evidence feature in policy documents is still lacking.

In this study, we aim to bridge this gap by analyzing the diverse range of evidence sources cited in policy documents authored by governments around the world. We ask: Whose knowledge—policy-based and scholarly—do governments rely on in their policy documents, and what do these patterns suggest about global dynamics of visibility and influence in evidence-informed policymaking? As policy documents increasingly serve as instruments for communicating decisions, justifying actions, and engaging with transnational audiences, the references they include offer a window into which forms of knowledge are recognized, legitimized, and shared across borders.

To investigate these questions, we draw on the Overton database—the largest online repository of publicly available policy texts—to examine patterns of citation across countries and policy domains. While this does not capture the full universe of knowledge used in policymaking, it reflects a particularly influential subset: the *documented*, *accessible*, and *digitally visible* evidence that circulates in the global policy arena. These documents are not only traceable and public-facing; they are also available to policy audiences beyond national borders, making them more likely to be discovered, consumed, and referenced by others. As such, they form part of the transnational infrastructure through which governments justify decisions, project legitimacy, and engage in global benchmarking. In this context, analyzing this corpus offers a window into how policy-relevant knowledge becomes visible, shared, and influential across the global policymaking landscape.

We engage in a systematic exercise of quantitative description, mapping citation patterns within this corpus of policy texts to understand whose knowledge becomes visible and authoritative in the documented landscape of global policymaking. Specifically, we examine whether policy documents tend to cite domestic versus foreign sources, identify which countries and scientific fields are most frequently referenced, and assess how these citation dynamics vary across geopolitical regions, economic stratification, institutional affiliations, and policy issue areas.

Our analyses reveal three key findings about the geography of evidence use. First, there are pronounced differences in the reliance on foreign versus domestic policy sources: governments in the Global South more frequently cite foreign materials, while higher-income countries tend to draw more heavily on domestic sources. Second, Global North countries dominate the production of both policy and academic research that is referenced in documents worldwide–with the United States playing an especially outsized role in shaping the global knowledge agenda. Third, although citation behavior varies across policy domains—with some relying more on academic knowledge or international references—the broader trend holds: evidence originating from economically powerful nations is consistently more prominent. These findings underscore persistent imbalances in knowledge circulation that feed into contemporary policymaking.

Empirical setting and contributions

To characterize the references used in policy documents, we integrate two large-scale data sources into a relational database, capturing information about government-authored documents and the materials they cite. To understand how governments draw on different forms of expertise, we distinguish between two major categories: *policy-based* sources, which include materials produced by policy actors such as governments, intergovernmental organizations (IGOs), and think tanks; and *scholarly* sources, which refer to scientific research outputs, including journal articles, working papers, and preprints authored by academic researchers.

Our dataset includes more than 1.2 million policy documents, citing 3.5 million scholarly works and 740,000 policy-based sources. Figure 1 presents an overview of our data collection structure, which forms the basis for our analyses. We extract the policy documents for our study from Overton, which compiles publicly accessible materials from over 1,700 policy sources using web-crawling techniques, gathering documents that are either written by policymakers or intended primarily for them (Szomszor and Adie, 2022). The database is largely language-agnostic, indexing and analyzing documents in multiple languages, making it a valuable resource for global policy analysis. Overton has become a key tool for studying how evidence informs policymaking, with prior research using it to track the integration of academic knowledge into policy discussions (Yin et al., 2021; Isett, Hicks, and Kingsley, 2025; Furnas, LaPira, and Wang, 2025; Ma and Cheng, 2025). Among available sources, Overton offers the most comprehensive coverage of policy documents, along with high accuracy in reference extraction (Szomszor and Adie, 2022; Jonkers et al., 2024).

Our study focuses on official government policy documents published between 2000 and 2024 that contain at least one reference to a scholarly or policy-based source. For each citation, we extract metadata on the referenced material, including its type and country of origin. While Overton provides structured information about both the citing policy documents and their policy-based references, its metadata on academic citations is more limited. To address this, we incorporate data from OpenAlex—an open, structured index of scholarly publications (Priem, Piwowar, and Orr, 2022). OpenAlex offers extensive metadata on academic works, including author affiliations, journal venues, topical classifications, and institutional locations. It has become a key infrastructure for metascientific research (Yu and Romero, 2024; Harris et al., 2024). By linking OpenAlex records to the academic citations found in Overton, we gain deeper insight into the geography and institutional sources of scholarly knowledge cited in policy. This enables us to enrich these references with deeper context—capturing not only what is cited, but also *who* produced it, *where* it was produced, and *in which domains* of scientific inquiry. Together, these two data sources allow us to capture a richer, granular view of evidence circulation in policy.

Our final database offers near-global geographical coverage spanning 185 countries. Importantly, the distribution of the core policy documents is not uniform. A substantially larger share of policy documents originates from Europe and the Americas, reflecting not only differences in publishing practices and institutional capacity, but also the nature of our empirical window: the *documented*, *accessible*, and *digitally visible* corpus of policy evidence. The availability of documents is highly correlated with the World Bank's Worldwide Governance Indicators (WGI) on regulatory quality, government effectiveness, and control of corruption (see *SI Appendix*, Fig A2). These patterns highlight the infrastructural and institutional factors around which knowledge becomes part of the visible layer of policy-relevant evidence. While this imbalance must be considered in interpreting global patterns, the breadth and scale of the dataset nonetheless offer a rare opportunity to analyze cross-national referencing practices within this influential domain of policy-relevant knowledge.

To structure our empirical analysis, we extract directed citation pairs between each policy document and the sources it cites. For policy-based references, we identify the authoring institution (e.g., government agency, think tank, or intergovernmental organization) and determine whether the source is domestic or foreign relative to the citing government. For academic references, we use OpenAlex metadata to associate each citation with the scientific discipline and national affiliation of the institutions of its authors.

To enable regional comparisons, we classify countries using the United Nations Regional Groups, which organize member states into five geopolitical clusters. We focus in particular on the Western European and Others Group (WEOG), which—together with observer states like the United States—closely aligns with conventional definitions of the Global North (Castro Torres and Alburez-Gutierrez, 2022; Zhou et al., 2022). We also draw on the UN M49 classification system, which distinguishes between 'developed' and 'developing' regions. These complementary schemes help us contextualize regional imbalances in both the production and citation of policy-relevant knowledge.

Our study contributes to research on the relationship between expert knowledge and policymaking in three key ways. First, we undertake a large-scale data collection effort centered on one of the most tangible and authoritative outputs of the policy process: official government documents. Unlike prior research, which has largely centered on the U.S. policy context, our dataset captures policy documents from a global perspective, enabling regional and cross-country comparisons. Second, we bring together two major categories of expert knowledge—policy-based sources and scientific research—bridging the gap between studies that have traditionally focused on one or the other. Third, we analyze evidence use across a wide range of policy domains, allowing us to examine how evidence use varies according to different fields with distinct expertise demands and 'localized' knowl-

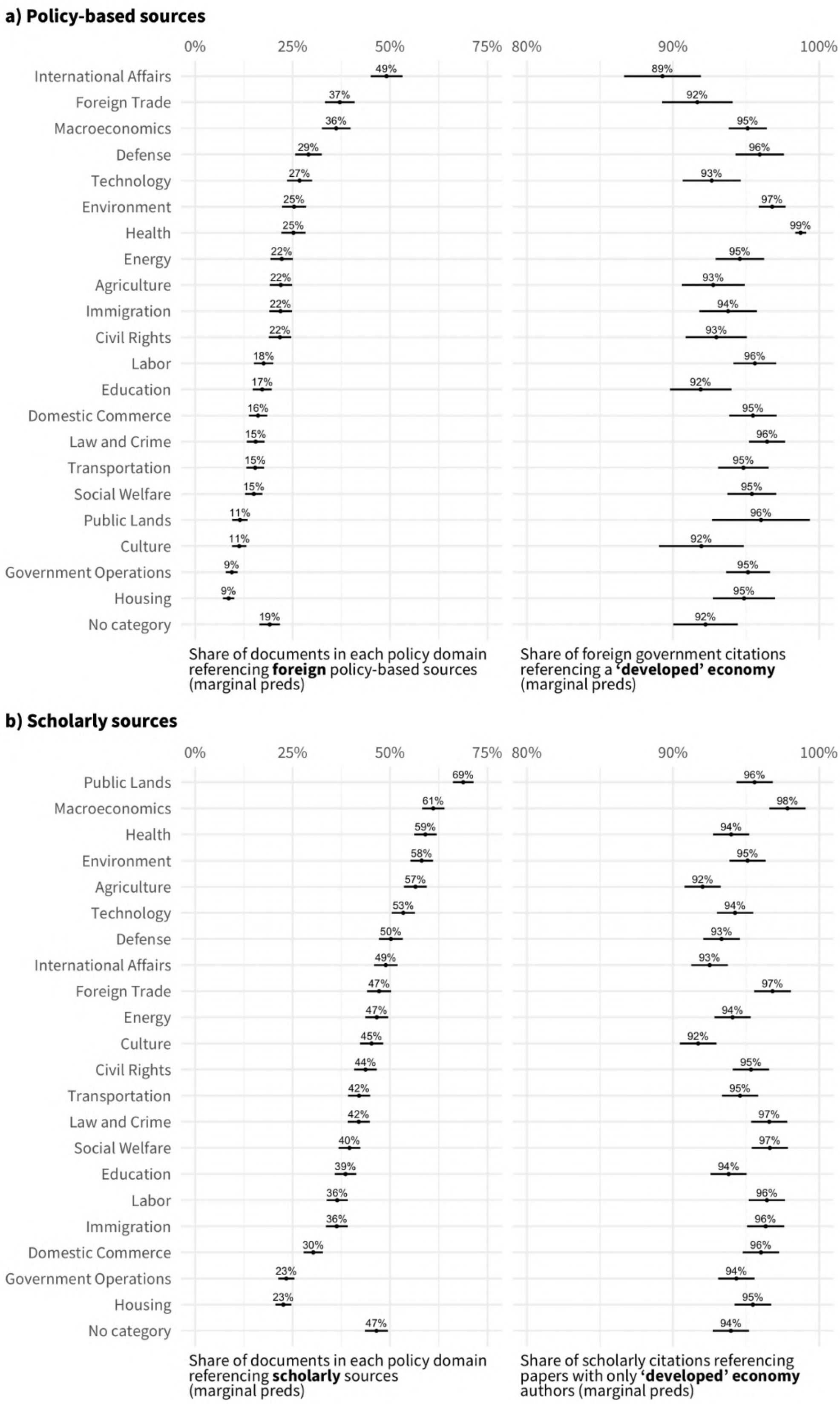


Figure 4: Overview of the distribution of references across policy domains. Results present the marginal predictions from logistic mixed-effects models with country of document random effects.

logistic regression models. The left-hand panels report estimates of the probability that a policy document within the respective policy domain references (a) a foreign policy-based source—defined as evidence from foreign governments, intergovernmental organizations (IGOs), or foreign think tanks—and (b) a scholarly source of any kind. These models are estimated over the full corpus of documents. The right-hand panels restrict the sample to documents that cite foreign government and scholarly sources, respectively, and model the probability that such citations originate from actors or authors based in 'developed' countries. Specifically, Panel (a) models whether a cited foreign government document comes from a 'developed' nation, and Panel (b) models whether cited academic papers have exclusively 'developed' country–based authors.

These models reveal domain-specific patterns. Engagement with foreign policy-based sources is highest in areas intrinsically tied to international coordination, such as International Affairs and Foreign Trade. In contrast, domains centered on domestic governance—like Cultural Affairs and Government Operations—rely far more on domestic sources. When it comes to scholarly references, we observe elevated engagement in technically intensive domains, including Public Lands, Macroeconomics, Health, Environment, Agriculture, and Technology. In these fields, more than half of the policy documents reference academic research, underscoring the importance of specialized expertise.

Among documents that include scholarly references, nearly 90% of citations fall within three broad disciplinary families: the medical, natural, and social sciences. While the total volume of citations from each of these families is relatively balanced, a domain-level breakdown reveals a consistent pattern: social science research is the most frequently cited in 16 of 21 policy domains.[3] These findings highlight not only the diversity of evidence practices across policy areas but also the central role of academic expertise—especially from the social sciences—in informing policy across a wide array of sectors.

Despite this domain-level variation in the type and source of evidence used, the right-hand panels of Figure 4 reveal a striking consistency: when policy documents cite foreign government sources or academic research, they overwhelmingly draw from knowledge produced in high-income countries. This pattern holds across nearly all policy domains. It suggests that while evidence needs may vary, the global geography of knowledge production—and the epistemic dependence on scholarship and policy knowledge from the Global North—remains a defining feature of contemporary evidence use in policymaking.

[3]The remaining domains—Public Lands, Agriculture, Environment, Health, and Energy—rely more heavily on the natural and medical sciences, reflecting their more technical orientation (see *SI Appendix, Figure B1*).

Discussion

This study offers a comprehensive quantitative account of evidence cited in government policy documents across 185 countries. By analyzing the world's largest corpus of *documented*, *accessible*, and *digitally visible* policy evidence, we gain insight into the information infrastructure that shapes policymaking across borders. Three broad empirical patterns stand out. First, we observe stark geographical differences in the sourcing of expert information. Governments in the Global North primarily cite domestic sources, whereas those in the Global South more often draw on foreign knowledge. Second, references to policy-based and academic research are heavily skewed toward work produced in the U.S., U.K., Canada, and other high-income countries. Even when diverse scientific disciplines contribute to policy discourse, the institutional and geographic origins of that knowledge remain tightly concentrated. Finally, we observe that while evidence needs vary by policy domain, the geographic origin of sources remains consistent: most knowledge, regardless of thematic focus, originates in the Global North.

Importantly, we do not aim to measure the direct impact of evidence on policymaking. Instead, we focus on the publicly visible layer of knowledge that circulates within policy documents which we argue holds important signals for understanding the flow of knowledge in the policy process. The citation patterns we uncover likely represent a dual reality: they reflect prevailing norms about what constitutes legitimate evidence in global policy discourse, and they illustrate the supply of expert knowledge that is available to policy professionals.

A related point to emphasize is that the data generation process is embedded in the very dynamics we seek to understand. Our dataset—and by extension, our findings—are embedded within the same infrastructure of visibility and access we aim to study. A large share of the documents in our corpus originate from Europe and the Americas. This is partly due to Overton's document indexing protocol, but more fundamentally, it reflects global differences in the capacity to produce, preserve, and disseminate policy materials. Rather than merely a methodological limitation, this constitutes an empirical finding in its own right: the availability of digitally visible policy evidence is unevenly distributed across the globe. Countries with more institutional capacity for public documentation are disproportionately represented in global policy repositories, which in turn reinforces their epistemic prominence. The infrastructure of evidence—who can publish, preserve, and make policy documents discoverable—can shape the profile of global knowledge hierarchies.

These asymmetries are further compounded by constraints on the demand side. While research shows that policy professionals tend to value evidence (Hjort et al., 2021; Senninger and Seeberg, 2024)—particularly when it is locally grounded (Vivalt, Coville, and Kc, 2025)—not all political actors are equally able to act on these preferences. Policymakers in low- and middle-income countries often face structural barriers such as limited research capacity and budgetary constraints. As a result, even where there is a desire for locally embedded evidence, decision-makers may be functionally reliant on internationally visible sources—most of which originate in high-income countries.

In fact, we find that countries most frequently cited as sources—such as the United States and the UK—are also among the world's largest investors in research and development relative to their GDP. These countries house robust academic ecosystems and dense bureaucratic apparatuses that generate a steady stream of policy-based evidence. However, the resilience of these knowledge systems should not be taken for granted. They remain politically contingent, subject to shifting funding priorities and ideological pressures. Recent developments indicate that political actors are increasingly willing to challenge the independence of scientific and policy institutions (Vandekerckhove et al., 2025; Garisto, Kozlov, and Tollefson, 2025). In the United States, for example, recent administrative actions have included downsizing parts of the federal policy workforce, reducing support for major research initiatives, and deprioritizing international collaboration—trends that may reduce the country's future visibility in global evidence flows. As a result, even countries that currently dominate the production and dissemination of policy-relevant knowledge may risk losing their status as global reference points, underscoring the fragility of epistemic centralization.

The implications of these asymmetries in expert knowledge circulation are not merely symbolic. Policy professionals grapple with complex challenges that can make organizations prone to getting stuck on local peaks—defaulting to a narrow set of familiar solutions rather than exploring a broader range of possibilities (Rivkin, 2000). When policymakers repeatedly draw on a narrow set of well-established sources, they may reinforce existing models and overlook alternatives that are more contextually appropriate or better suited to emerging problems. In contrast, incorporating a broader range of knowledge can expand the space of viable solutions. From a systems perspective, some heterogeneity in evidence use can help prevent informational lock-in and foster the exploration of new policy pathways (Gomez and Lazer, 2019). This diversity is not only about inclusion; it plays a crucial role in strengthening resilience, enabling innovation, and enhancing the global policy community's ability to respond to complex problems.

Taken together, our findings underscore how global policymaking is shaped not only by ideas, but by infrastructures of visibility, access, and epistemic legitimacy. The expert knowledge that circulates in official policy documents is not evenly distributed—it reflects entrenched relationships in who produces policy-relevant evidence, who gets cited, and who gets seen.

Materials and methods

Data collection

The data on policy documents were sourced from Overton, the largest database of policy documents, which indexes over 44,000 sources from governments, intergovernmental organizations, and think tanks (Szomszor and Adie, 2022). We used the Overton API to extract data for all available documents authored by government sources published within the 25-year range from 2000 to 2024. The data collection was conducted between January 17 to 23, 2025.

In total, we compiled 1,243,768 policy documents from 185 countries. To enrich this collection, we also extracted data on all entities cited by these documents, with a particular focus on policy documents and academic works. This resulted in 740,742 unique policy documents—including sources outside government—and 3,518,012 academic works, along with their respective DOIs.

To contextualize the academic research referenced in the policy documents, we extracted auxiliary information from OpenAlex, an open index of academic work (Priem, Piwowar, and Orr, 2022). This auxiliary data includes metadata about each academic work, such as authorship, citation metrics, and publication details.

Measurement

Policy domains. We categorize policy documents into specific policy domains. We approached this task as an automated text-annotation problem using a large language model (LLM). Specifically, we classified the policy domains of individual policy documents based on the framework established by the Comparative Agendas Project (CAP) (Baumgartner, Breunig, and Grossman, 2019). Our classification process involved assigning each policy document summary provided by Overton to one of the predefined CAP categories.

Following the recommendations by Barrie, Palmer, and Spirling (2025), we favored a locally versioned open source language model, over similarly performing commercial alternatives (e.g., OpenAI's GPT). Our classification was performed using Meta's Llama3.3-70B model, which was prompted with a standardized annotation instruction. Each policy document was classified in its own session to prevent prior classifications from influencing subsequent inferences.

To ensure the validity and reliability of our automated classification, we benchmarked the Llama3.3-70B categorization against a human annotator on a random draw of 1,000 policy documents. The LLM derived categorization coincided with the annotator's in 78.9% of the instances. This level of agreement is comparable to the reported intercoder reliability of the CAP in the Croatian, UK, and German legislative contexts (Baumgartner, Breunig, and Grossman, 2019). Notably, the zero-shot performance of the Llama3.3-70B model exceeded that of the CAP Babel Machine, an encoder model employing XLM-RoBERTa trained for CAP classification (Sebők et al., 2024), which was used as an additional benchmark (see SI Appendix A.3 for a detailed description of the protocol).

Auxiliary variables

To assess the international reach of policy documents, we calculate a series of metrics from the from the government-to-government citation matrix. We utilize a simple *H*-index focused on total citations per documents, such that: $h = \max\{h' \mid \text{at least } h' \text{ papers have } \geq h' \text{ citations}\}$. Under the same logic, we extract a country-centered *H*-index, which instead of counting total citations per document, focuses how many different countries have cited a given country's documents. Finally, we extract the betweenness, eigenvector, and PageRank centrality measures from the citation network. For our supplementary analysis of citations to bordering countries, we extracted contiguity relationships between states from the CoW Direct Contiguity (v3.2) dataset (Stinnett et al., 2002).

Acknowledgments

We thank Asya Magazinnik, Will Lowe, Jesper A. Hansen, Simon Munzert, Francesca Minetto, Sascha Göbel, as well as audiences at the Data Science Lab Research Group, University of Lucerne, and EPSA, for their valuable feedback. SRR acknowledges financial support by the Hertie School in the form of a doctoral stipend. RS acknowledges funding from the Carlsberg Foundation (Grant No. CF22-0805).

References

African Union (2021). *Accurate and Disaggregated Data Critical for Evidence-Based Policies: Launch of Africa Migration Data Network*. Press Release No: /2021, Venue: Addis Ababa, Ethiopia.

Bachrach, Peter and Morton S. Baratz (1962). "Two Faces of Power". *American Political Science Review* 56.4, pp. 947–952.

Barrie, Christopher, Alexis Palmer, and Arthur Spirling (2025). *Replication for Language Models: Problems, Principles, and Best Practice for Political Science*.

Baumgartner, Frank R., Christian Breunig, and Emiliano Grossman, eds. (2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press.

Baumgartner, Frank R. and Bryan D. Jones (2015). *The Politics of Information: Problem Definition and the Course of Public Policy in America*. Chicago: University of Chicago Press.

Bornmann, Lutz, Robin Haunschild, and Werner Marx (2016). "Policy documents as sources for measuring societal impact: how often is climate change research mentioned in policy-related documents?" *Scientometrics* 109.3, pp. 1477–1495.

Campbell, John L. (2002). "Ideas, Politics, and Public Policy". *Annual Review of Sociology* 28, pp. 21–38.

Campbell, John L. and Ove K. Pedersen (2014). *The National Origins of Policy Ideas: Knowledge Regimes in the United States, France, Germany, and Denmark*. Princeton, NJ: Princeton University Press.

Castro Torres, Andrés F. and Diego Alburez-Gutierrez (2022). "North and South: Naming practices and the hidden dimension of global disparities in knowledge production". *Proceedings of the National Academy of Sciences* 119.10, e2119373119.

Commission, European (2014). *Mission letter to Commissioner Moedas*.

Freeman, Richard and Jo Maybin (2011). "Documents, practices and policy".

Furnas, Alexander C., Timothy M. LaPira, and Dashun Wang (2025). "Partisan disparities in the use of science in policy". *Science* 388.6745, pp. 362–367.

Garisto, Dan, Max Kozlov, and Jeff Tollefson (2025). "Trump team orders huge government lay-offs: how science could fare". *Nature*.

Gomez, Charles J., Andrew C. Herman, and Paolo Parigi (2022). "Leading countries in global science increasingly receive more citations than other countries doing similar research". *Nature Human Behaviour* 6.7, pp. 919–929.

Gomez, Charles J. and David M. J. Lazer (2019). "Clustering knowledge and dispersing abilities enhances collective problem solving in a network". *Nature Communications* 10.1, p. 5146.

Harris, Mallory J, Ryan Murtfeldt, Shufan Wang, Erin A Mordecai, and Jevin D West (2024). "Perceived experts are prevalent and influential within an antivaccine community on Twitter". *PNAS Nexus* 3.2, pgae007.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini (2021). "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities". *American Economic Review* 111.5, pp. 1442–1480.

Isett, Kimberley R., Diana Hicks, and Gordon Kingsley (2025). "Blending the two communities: concurrent evidence use in state-level policy processes". *Public Management Review* 0.0, pp. 1–18.

Jones, Bryan D. and Frank R. Baumgartner (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago: University of Chicago Press.

Jonkers, Koen, DANTAS FARIA Jorge Costa, Federico Biagi, and Mario Scharfbillig (2024). *Quantitative "science for policy" evaluation: are Science for Policy Organisations operating in Pasteur's Science for Policy Quadrant?*

Kenny, Caroline, David Christian Rose, Abbi Hobbs, Chris Tyler, and Jason Blackstock (2017). *The role of research in the UK Parliament*. Vol. 1. Parliamentary Office of Science and Technology.

Kingdon, John W. (1984). *Agendas, alternatives, and public policies*. S.l.: Harper Collins.

Kuhlmann, Johanna, Delia González de Reufels, Klaus Schlichte, and Frank Nullmeier (2020). "How social policy travels: A refined model of diffusion". *Global Social Policy* 20.1, pp. 80–96.

Lindblom, Charles E. (1959). "The Science of "Muddling Through"". *Public Administration Review* 19.2, pp. 79–88.

Lorenc, Theo, Elizabeth F. Tyner, Mark Petticrew, Steven Duffy, Fred P. Martineau, Gemma Phillips, and Karen Lock (2014). "Cultures of evidence across policy sectors: systematic review of qualitative evidence". *European Journal of Public Health* 24.6, pp. 1041–1047.

Ma, Ji and Yuan Daniel Cheng (2025). "Why do some academic articles receive more citations from policy communities?" *Public Administration Review*.

National Research Council (2012). *Using Science as Evidence in Public Policy*. Washington, D.C.: National Academies Press.
Nielsen, Mathias Wullum and Jens Peter Andersen (2021). “Global citation inequality is on the rise”. *Proceedings of the National Academy of Sciences* 118.7, e2012208118.
Priem, Jason, Heather Piwowar, and Richard Orr (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*.
Rivkin, Jan W. (2000). “Imitation of Complex Strategies”. *Management Science* 46.6, pp. 824–844.
Schattschneider, E. E. (1960). *The Semi-Sovereign People: A Realist’s View of Democracy in America*. New York: Holt, Rinehart and Winston.
Sebők, Miklós, Ákos Máté, Orsolya Ring, Viktor Kovács, and Richárd Lehoczki (2024). “Leveraging Open Large Language Models for Multilingual Policy Topic Classification: The Babel Machine Approach”. *Social Science Computer Review*, p. 08944393241259434.
Senninger, Roman and Henrik B. Seeberg (2024). “Which Information Do Politicians Pay Attention To? Evidence from a Field Experiment and Interviews”. *British Journal of Political Science*, pp. 1–18.
Simon, Herbert A. (1944). “Decision-Making and Administrative Organization”. *Public Administration Review* 4.1, p. 16.
Stinnett, Douglas M., Jaroslav Tir, Paul F. Diehl, Philip Schafer, and Charles Gochman (2002). “The Correlates of War (Cow) Project Direct Contiguity Data, Version 3.0”. *Conflict Management and Peace Science* 19.2, pp. 59–67.
Szomszor, Martin and Euan Adie (2022). “Overton: A bibliometric database of policy document citations”. *Quantitative Science Studies* 3.3, pp. 624–650.
United States Congress (2019). *Foundations for Evidence-Based Policymaking Act*.
Vandekerckhove, Joachim, David Kellen, Jennifer S. Trueblood, and Richard M. Shiffrin (2025). *The structures that fund, publish, and evaluate science shape its progress*. *PNAS Updates*, Proceedings of the National Academy of Sciences.
Vivalt, Eva, Aidan Coville, and Sampada Kc (2025). “Local knowledge, formal evidence, and policy decisions”. *Journal of Development Economics* 174, p. 103425.
Weible, Christopher M. (2008). “Expert-Based Information and Policy Subsystems: A Review and Synthesis”. *Policy Studies Journal* 36.4, pp. 615–635.
Yin, Yian, Jian Gao, Benjamin F Jones, and Dashun Wang (2021). “Coevolution of policy and science during the pandemic”. *Science* 371.6525, pp. 128–130.
Yu, Yulin and Daniel M. Romero (2024). “Does the use of unusual combinations of datasets contribute to greater scientific impact?” *Proceedings of the National Academy of Sciences* 121.41, e2402802121.
Zhou, Yuyu, Xuecao Li, Wei Chen, Lin Meng, Qiusheng Wu, Peng Gong, and Karen C. Seto (2022). “Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the Global South”. *Proceedings of the National Academy of Sciences* 119.46, e2214813119.

Table A1: Overview of core document data table.

Country	Group	# Docs	Country	Group	# Docs	Country	Group	# Docs
USA	WEOG	558925	Georgia	EEG	582	Malawi	AF-G	75
UK	WEOG	110409	Trinidad and Tobago	GRULAC	573	Nicaragua	GRULAC	73
Spain	WEOG	77886	Cyprus	APG	552	Armenia	EEG	65
Sweden	WEOG	75114	Ethiopia	AF-G	486	Haiti	GRULAC	65
Canada	WEOG	71818	Sri Lanka	APG	404	Benin	AF-G	64
Australia	WEOG	41726	Dominican Republic	GRULAC	396	Sierra Leone	AF-G	58
Germany	WEOG	28848	Jamaica	GRULAC	379	Senegal	AF-G	57
Indonesia	APG	20715	Kosovo	EEG*	377	Kiribati	APG	55
Japan	APG	20460	Nepal	APG	360	Suriname	GRULAC	54
France	WEOG	18194	Barbados	GRULAC	358	Belarus	EEG	52
Switzerland	WEOG	16763	Maldives	APG	358	Cuba	GRULAC	51
Netherlands	WEOG	15921	Lebanon	APG	353	Vanuatu	APG	50
Finland	WEOG	13690	Kenya	AF-G	346	UAE	APG	49
Brazil	GRULAC	12860	Morocco	AF-G	343	Solomon Islands	APG	46
Portugal	WEOG	11828	Namibia	AF-G	342	Eswatini	AF-G	44
Ireland	WEOG	10653	El Salvador	GRULAC	336	Micronesia	APG	40
Belgium	WEOG	10409	Guatemala	GRULAC	334	Sudan	AF-G	39
Austria	WEOG	9932	Croatia	EEG	321	Cape Verde	AF-G	38
Mexico	GRULAC	7987	Rwanda	AF-G	307	Jordan	APG	37
New Zealand	WEOG	7352	Nigeria	AF-G	262	Timor Leste	APG	35
Peru	GRULAC	6644	Bolivia	GRULAC	254	Seychelles	AF-G	34
Italy	WEOG	6134	Ghana	AF-G	252	Bahrain	APG	33
Colombia	GRULAC	5940	Honduras	GRULAC	230	Iraq	APG	32
Norway	WEOG	5754	Vietnam	APG	223	Mongolia	APG	31
Turkey	WEOG	5295	Pakistan	APG	214	Tonga	APG	30
Chile	GRULAC	4015	Panama	GRULAC	201	Zimbabwe	AF-G	30
Argentina	GRULAC	3788	Bulgaria	EEG	192	Angola	AF-G	28
Latvia	EEG	3508	Afghanistan	APG	186	Lesotho	AF-G	28
Singapore	APG	3478	Belize	GRULAC	178	Mauritania	AF-G	28
Uruguay	GRULAC	3310	Saudi Arabia	APG	178	Myanmar	APG	23
South Africa	AF-G	3235	Bahamas	GRULAC	169	Togo	AF-G	22
Slovenia	EEG	2248	Bhutan	APG	163	San Marino	WEOG	21
Denmark	WEOG	2202	Malta	WEOG	161	Syria	APG	21
Estonia	EEG	2053	Bos and Herz	EEG	160	Ivory Coast	AF-G	20
Czech Republic	EEG	1868	Cameroon	AF-G	160	Saint Kitts and Nevis	GRULAC	20
Romania	EEG	1814	Albania	EEG	154	Tajikistan	APG	18
China	APG	1726	Russia	EEG	152	Liechtenstein	WEOG	17
Poland	EEG	1590	Mauritius	AF-G	143	Mali	AF-G	17
India	APG	1546	Papua New Guinea	APG	139	Burundi	AF-G	16
Philippines	APG	1416	Oman	APG	137	Palestine	APG	16
Taiwan	APG*	1273	Liberia	AF-G	133	DR Congo	AF-G	15
Lithuania	EEG	1212	Palau	APG	132	Uzbekistan	APG	15
Uganda	AF-G	1071	Montenegro	EEG	130	Kuwait	APG	13
North Macedonia	EEG	1064	Cambodia	APG	127	Sao Tome and Principe	AF-G	13
Malaysia	APG	1044	Tunisia	AF-G	125	Marshall Islands	APG	10
Thailand	APG	1000	Andorra	WEOG	123	Niger	AF-G	9
Ukraine	EEG	1000	Kyrgyzstan	APG	123	Algeria	AF-G	8
Ecuador	GRULAC	993	Venezuela	GRULAC	123	Gambia	AF-G	8
Tanzania	AF-G	981	Qatar	APG	116	CAR	AF-G	7
Slovakia	EEG	951	Guyana	GRULAC	112	Chad	AF-G	7
South Korea	APG	881	Fiji	APG	102	Yemen	APG	4
Iceland	WEOG	853	Mozambique	AF-G	95	Djibouti	AF-G	3
Costa Rica	GRULAC	808	Samoa	APG	94	Madagascar	AF-G	3
Moldova	EEG	765	Zambia	AF-G	94	South Sudan	AF-G	3
Luxembourg	WEOG	751	Brunei	APG	90	Equatorial Guinea	AF-G	2
Azerbaijan	EEG	725	Bangladesh	APG	89	Eritrea	AF-G	2
Paraguay	GRULAC	720	Guinea	AF-G	89	Nauru	APG	2
Hungary	EEG	681	Burkina Faso	AF-G	87	Comoros	AF-G	1
Serbia	EEG	677	Laos	APG	84	Monaco	WEOG	1
Greece	WEOG	655	Botswana	AF-G	83	Turkmenistan	APG	1
Iran	APG	637	Egypt	AF-G	80	Vatican City	WEOG	1
Israel	WEOG	588	Somalia	AF-G	78			

A.2 Government-to-government reference reach

Table A2: H-index and Country-centered H-index of government policy documents by country.

Country	H-i	CH-i	# Refs	Country	H-i	CH-i	# Refs	Country	H-i	CH-i	# Refs	Country	H-i	CH-i	# Refs
USA	44	19	10553	Rwanda	3	2	15	Trinidad and Tobago	2	1	17	Kosovo	1	1	10
UK	32	15	10508	Singapore	3	2	182	Turkey	2	1	29	Lebanon	1	1	3
Germany	12	5	986	South Africa	3	2	47	Uganda	2	1	20	Lesotho	1	1	1
Australia	11	6	1056	Afghanistan	2	2	12	United Arab Emirates	2	2	3	Mali	1	1	3
Canada	11	7	1596	Bahamas	2	1	15	Albania	1	1	1	Marshall Islands	1	1	2
Ireland	11	4	978	Bhutan	2	2	5	Algeria	1	1	3	Mauritania	1	1	1
Netherlands	10	5	663	Costa Rica	2	2	28	Andorra	1	1	2	Mauritius	1	1	6
Norway	9	4	495	Croatia	2	2	12	Angola	1	1	1	Moldova	1	1	3
France	8	3	706	Ecuador	2	2	20	Armenia	1	1	5	Mongolia	1	1	2
India	8	5	207	El Salvador	2	1	12	Azerbaijan	1	1	3	Montenegro	1	1	4
New Zealand	8	5	533	Ghana	2	2	22	Bahrain	1	1	2	Mozambique	1	1	4
Sweden	8	4	967	Greece	2	1	5	Bangladesh	1	1	5	Myanmar	1	1	2
Switzerland	8	5	411	Guatemala	2	2	14	Barbados	1	1	14	Niger	1	1	1
Chile	7	4	171	Haiti	2	1	5	Belize	1	1	5	Oman	1	1	1
Finland	7	4	502	Iceland	2	2	18	Benin	1	1	2	Palau	1	1	12
Japan	7	5	494	Kenya	2	1	19	Bolivia	1	1	12	Qatar	1	1	6
Spain	7	5	628	Latvia	2	1	27	Bosnia and Herzegovina	1	1	3	Russia	1	1	1
Belgium	6	4	412	Liberia	2	2	8	Botswana	1	1	4	Saint Kitts and Nevis	1	1	3
Colombia	6	4	283	Luxembourg	2	2	14	Brunei	1	1	2	Samoa	1	1	6
Denmark	6	3	129	Malawi	2	1	6	Burkina Faso	1	1	1	San Marino	1	1	1
Austria	5	3	176	Malaysia	2	1	9	Burundi	1	1	2	Sao Tome and Principe	1	1	1
China	5	3	65	Maldives	2	1	4	Cambodia	1	1	2	Saudi Arabia	1	1	4
Uruguay	5	2	79	Malta	2	2	16	Cameroon	1	1	8	Sierra Leone	1	1	7
Argentina	4	2	71	Morocco	2	1	15	Cape Verde	1	1	2	South Sudan	1	1	2
Brazil	4	4	122	Namibia	2	1	4	Cuba	1	1	1	Sri Lanka	1	1	11
Czech Republic	4	2	35	Nepal	2	1	19	Cyprus	1	1	6	Sudan	1	1	3
Estonia	4	3	65	Nicaragua	2	1	7	Dominican Republic	1	1	10	Taiwan	1	1	20
Italy	4	2	157	Nigeria	2	2	25	Egypt	1	1	5	Togo	1	1	1
Mexico	4	3	130	North Macedonia	2	1	8	Eritrea	1	1	1	Tunisia	1	1	4
Paraguay	4	1	12	Panama	2	1	9	Eswatini	1	1	3	Ukraine	1	1	4
Portugal	4	3	81	Papua New Guinea	2	2	18	Ethiopia	1	1	3	Vanuatu	1	1	4
Bulgaria	3	1	8	Romania	2	1	13	Fiji	1	1	6	Venezuela	1	1	2
Hungary	3	2	10	Senegal	2	2	10	Georgia	1	1	8	Vietnam	1	1	2
Indonesia	3	1	16	Serbia	2	2	8	Guinea	1	1	1	Zambia	1	1	5
Jamaica	3	2	33	Slovakia	2	1	22	Guyana	1	1	6				
Lithuania	3	3	38	Slovenia	2	2	22	Honduras	1	1	1				
Pakistan	3	2	18	South Korea	2	1	6	Iran	1	1	2				
Peru	3	3	74	Tanzania	2	2	57	Ivory Coast	1	1	6				
Philippines	3	1	32	Thailand	2	1	19	Jordan	1	1	3				
Poland	3	2	16	Timor Leste	2	1	9	Kiribati	1	1	3				

Table A3: Inverse document frequency weighted governent-to-government citations.

H-i rank	Country	IFW refs	H-i rank	Country	IFW refs	H-i rank	Country	IFW refs	H-i rank	Country	IFW refs
2	UK	35312	37	Pakistan	161	61	Malaysia	63	111	Ethiopia	25
1	USA	12714	57	Latvia	156	67	Nicaragua	61	90	Azerbaijan	24
6	Ireland	5278	71	Papua New Guinea	152	118	Ivory Coast	60	97	Bosnia and Herzegovina	23
5	Canada	5057	66	Nepal	151	60	Malawi	59	122	Lebanon	21
4	Australia	4067	56	Kenya	150	69	North Macedonia	56	128	Moldova	21
3	Germany	3665	75	Slovakia	140	112	Fiji	56	144	South Sudan	21
12	Sweden	3640	83	Uganda	139	113	Georgia	54	101	Burundi	20
11	New Zealand	3211	49	Ecuador	137	74	Serbia	53	125	Marshall Islands	20
7	Netherlands	3025	41	Rwanda	136	136	Qatar	53	152	Venezuela	19
8	Norway	2981	63	Malta	136	115	Guyana	51	87	Andorra	18
9	France	2758	76	Slovenia	135	139	Samoa	51	91	Bahrain	18
15	Finland	2724	82	Turkey	135	32	Bulgaria	49	95	Benin	17
18	Belgium	2182	45	Bahamas	118	77	South Korea	46	104	Cape Verde	17
13	Switzerland	1974	55	Iceland	115	94	Belize	45	129	Mongolia	17
17	Spain	1625	81	Trinidad and Tobago	115	127	Mauritius	45	132	Myanmar	17
19	Colombia	1482	79	Thailand	109	89	Armenia	44	99	Brunei	16
16	Japan	1375	135	Palau	107	54	Haiti	43	102	Cambodia	16
10	India	1201	93	Barbados	106	154	Zambia	43	117	Iran	15
42	Singapore	970	40	Poland	102	46	Bhutan	41	153	Vietnam	15
14	Chile	939	147	Taiwan	101	106	Cyprus	40	109	Eritrea	13
21	Austria	834	64	Morocco	99	92	Bangladesh	39	123	Lesotho	11
20	Denmark	739	73	Senegal	99	108	Egypt	38	88	Angola	10
28	Italy	705	59	Luxembourg	94	151	Vanuatu	38	114	Guinea	10
29	Mexico	643	96	Bolivia	94	98	Botswana	36	133	Niger	10
25	Brazil	489	53	Guatemala	92	131	Mozambique	35	140	San Marino	10
27	Estonia	436	44	Afghanistan	91	138	Saint Kitts and Nevis	35	141	Sao Tome and Principe	10
23	Uruguay	402	48	Croatia	90	142	Saudi Arabia	35	100	Burkina Faso	9
24	Argentina	401	50	El Salvador	87	65	Namibia	34	105	Cuba	9
31	Portugal	373	34	Indonesia	81	52	Greece	33	126	Mauritania	9
38	Peru	347	30	Paraguay	78	62	Maldives	32	148	Togo	9
22	China	321	145	Sri Lanka	78	130	Montenegro	32	85	Albania	8
78	Tanzania	320	80	Timor Leste	75	146	Sudan	30	134	Oman	8
35	Jamaica	256	107	Dominican Republic	73	149	Tunisia	30	116	Honduras	7
43	South Africa	241	143	Sierra Leone	72	84	United Arab Emirates	28	137	Russia	7
36	Lithuania	236	121	Kosovo	71	110	Eswatini	28			
68	Nigeria	217	70	Panama	67	124	Mali	28			
47	Costa Rica	216	72	Romania	67	86	Algeria	27			
26	Czech Republic	188	103	Cameroon	66	119	Jordan	27			
51	Ghana	182	33	Hungary	65	150	Ukraine	27			
39	Philippines	169	58	Liberia	65	120	Kiribati	26			

Note: We generate the weights based on the total number of available documents from each country in the larger Overton pool, not just the subset of 1.2 million documents that engage in referencing. The weight is calculated by $\log\left(\frac{1}{\text{Document Frequency of Country X}} + 1\right)$

Table A4: Descriptives and centrality measures from government-to-government citation network.

Country	In-D	Out-D	BC	EVC	PR	Country	In-D	Out-D	BC	EVC	PR
United States	11172	97438	15324.176	0.554	0.083	Bhutan	39	14	0.012	0.001	0.001
Spain	19697	1079	975.219	1	0.08	Papua New Guinea	30	27	0.033	0.002	0.001
United Kingdom	13739	41842	4228.52	0.651	0.044	Namibia	34	6	0.002	0.002	0.001
Germany	8280	6445	1288.349	0.409	0.04	Zambia	32	5	0	0.002	0.001
Sweden	11952	2604	215.375	0.582	0.037	Vanuatu	5	4	0	0	0.001
Mexico	2245	408	42.912	0.123	0.037	Malawi	34	9	0.001	0.002	0.001
Canada	18766	4691	218.27	0.981	0.03	Saudi Arabia	103	9	0.001	0.006	0.001
Australia	9855	2782	302.329	0.515	0.027	Luxembourg	133	20	0.001	0.006	0.001
Switzerland	4886	1754	798.109	0.245	0.023	Senegal	47	56	0.689	0.002	0.001
France	5865	9817	1833.856	0.312	0.022	Egypt	144	51	0.437	0.007	0.001
Finland	5486	1010	18.072	0.286	0.022	Ghana	58	29	0.051	0.003	0.001
Italy	4611	546	272.132	0.228	0.021	Kuwait	9	4	0	0	0.001
Japan	5738	830	426.715	0.296	0.021	St. Kitts & Nevis	18	3	0.005	0.001	0.001
Argentina	1557	201	13.163	0.071	0.021	Guyana	18	7	0.001	0.001	0.001
Netherlands	6661	2050	133.782	0.33	0.02	South Korea	204	19	0.003	0.011	0.001
Peru	2468	121	4.8	0.121	0.02	Lesotho	14	1	0	0.001	0.001
Norway	4162	1395	238.988	0.212	0.019	Sudan	17	6	0	0.001	0.001
Portugal	2908	174	46.744	0.147	0.019	Bahamas	49	17	0.008	0.002	0.001
Uruguay	1196	145	2.19	0.06	0.018	Nigeria	99	36	0.065	0.005	0.001
Chile	2300	390	186.619	0.123	0.017	Armenia	69	6	0	0.004	0.001
Ireland	5337	2813	172.09	0.297	0.015	San Marino	14	1	0	0.001	0.001
Belgium	4129	6268	364.291	0.21	0.014	Kiribati	20	3	0.005	0.001	0.001
Austria	2442	880	88.087	0.104	0.014	Morocco	49	26	65.045	0.003	0.001
Colombia	2828	453	30.037	0.147	0.014	Bosnia & Herz	84	5	0.406	0.004	0.001
Turkey	2162	58	20.62	0.106	0.012	Qatar	210	37	0	0.012	0.001
Brazil	2141	237	13.175	0.105	0.011	UAE	71	6	2.087	0.003	0.001
New Zealand	2833	977	9.602	0.153	0.008	Albania	101	2	0.005	0.005	0.001
Latvia	1264	35	1.09	0.061	0.008	Lebanon	118	8	0	0.007	0.001
Romania	1385	17	2.014	0.066	0.008	Sierra Leone	36	8	0.001	0.002	0.001
Moldova	811	3	2.809	0.038	0.008	Liberia	27	10	0	0.001	0.001
Costa Rica	209	43	0.281	0.011	0.007	Tuvalu	0	3	0	0	0.001
Slovenia	887	29	2.393	0.042	0.006	Tunisia	46	4	0	0.002	0.001
China	943	249	40.713	0.047	0.006	Benin	19	2	0	0.001	0.001
Indonesia	797	37	44.307	0.038	0.006	Togo	4	1	0	0	0.001
Denmark	888	589	2.074	0.045	0.005	North Macedonia	441	11	0.001	0.023	0.001
Lithuania	356	89	19.471	0.016	0.005	Bulgaria	51	32	0.001	0.003	0.001
Uganda	1219	39	0.162	0.061	0.005	Cameroon	71	8	0.078	0.004	0.001
Estonia	756	182	4.478	0.037	0.005	Burundi	10	2	0	0	0.001
Greece	829	23	56.527	0.045	0.005	Georgia	313	11	0.251	0.015	0.001
Dominican Republic	281	13	1.6	0.013	0.005	Ethiopia	158	3	0	0.008	0.001
Paraguay	222	62	0.032	0.01	0.004	Eritrea	0	2	0	0	0.001
South Africa	1751	218	1.671	0.094	0.004	Vietnam	14	4	0	0.001	0.001
Ecuador	138	28	0.075	0.007	0.004	Côte d'Ivoire	4	7	0	0	0.001
Bolivia	110	15	0.011	0.004	0.004	Jordan	8	3	0	0	0.001
Singapore	1118	442	2.133	0.058	0.004	Somalia	21	3	0	0.001	0.001
Hungary	303	26	1.457	0.013	0.004	Algeria	1	4	0	0	0.001
Trinidad & Tobago	281	19	0.176	0.013	0.004	Myanmar (Burma)	10	2	0	0	0.001
Barbados	160	15	195.134	0.007	0.004	Mozambique	31	4	0	0.001	0.001
Israel	475	46	0.815	0.025	0.003	Azerbaijan	20	3	0	0.001	0.001
Sri Lanka	800	14	0.319	0.04	0.003	Maldives	67	9	0	0.003	0.001
Philippines	548	56	11.517	0.027	0.003	Brunei	50	2	0.001	0.002	0.001
Czechia	501	90	1.559	0.025	0.003	Palau	19	13	0	0.001	0.001
Croatia	138	20	117.59	0.007	0.003	Iran	69	8	0	0.004	0.001
Tanzania	325	97	1.055	0.015	0.003	Cambodia	62	2	0.001	0.003	0.001
Kyrgyzstan	50	0	0	0.003	0.003	South Sudan	0	3	0	0	0.001
Rwanda	165	26	0.409	0.008	0.002	Eswatini	71	5	0	0.004	0.001
Venezuela	81	148	0.742	0.004	0.002	Bahrain	4	3	0.001	0	0.001
India	407	789	3.734	0.02	0.002	Cape Verde	2	2	0	0	0.001
Russia	162	121	0.358	0.008	0.002	Marshall Islands	0	2	0	0	0.001
Panama	34	12	0.064	0.001	0.002	Niger	2	1	0	0	0.001
Thailand	291	95	1.45	0.016	0.002	Mongolia	7	9	0	0	0.001
Bangladesh	39	9	0.227	0.002	0.002	Oman	16	2	0	0.001	0.001
Timor-Leste	19	13	0.003	0.001	0.002	Angola	12	1	0	0	0.001
Nepal	204	22	0.192	0.01	0.002	Palestinian Territ	4	2	0	0	0.001
Poland	202	88	0.015	0.009	0.002	Libya	0	1	0	0	0.001
Malta	58	26	0.054	0.003	0.002	Mali	12	3	0	0.001	0.001
Montenegro	41	5	0.123	0.002	0.002	São Tomé & Príncipe	1	1	0	0	0.001
Slovakia	671	53	0.471	0.035	0.002	Guinea	47	1	0	0.002	0.001
Jamaica	189	48	1.389	0.01	0.002	Mauritania	2	2	0	0	0.001
Pakistan	134	47	0.218	0.007	0.002	Burkina Faso	4	1	0	0	0.001
Belize	126	5	36.901	0.006	0.002	Belarus	17	0	0	0.001	0.001
Guatemala	173	23	0.016	0.009	0.002	Chad	1	0	0	0	0.001
Serbia	213	14	0.087	0.011	0.002	DR Congo	7	0	0	0	0.001
Botswana	55	5	172.001	0.002	0.002	Gambia	2	0	0	0	0.001
Malaysia	848	12	0.013	0.046	0.002	Iraq	49	0	0	0.002	0.001
Iceland	331	39	0.067	0.017	0.002	Kosovo	108	0	0	0.006	0.001
Nicaragua	8	11	0.013	0	0.002	Laos	12	0	0	0	0.001
Mauritius	118	6	0.098	0.006	0.002	Liechtenstein	7	0	0	0	0.001
El Salvador	78	18	0.203	0.004	0.002	Madagascar	2	0	0	0	0.001
Cyprus	144	11	0.004	0.006	0.002	Micronesia	17	0	0	0.001	0.001
Taiwan	429	20	0.026	0.022	0.002	Monaco	1	0	0	0	0.001
Ukraine	99	8	0.01	0.005	0.002	Seychelles	6	0	0	0	0.001
Cuba	98	1	0	0.005	0.002	Solomon Islands	12	0	0	0.001	0.001
Andorra	36	2	0	0.002	0.002	Suriname	13	0	0	0.001	0.001
Honduras	29	1	0	0.001	0.002	Syria	7	0	0	0	0.001
Haiti	16	8	0.003	0.001	0.001	Tajikistan	2	0	0	0	0.001
Kenya	107	123	0.212	0.005	0.001	Tonga	11	0	0	0.001	0.001
Fiji	12	23	0.007	0.001	0.001	Turkmenistan	3	0	0	0	0.001
Afghanistan	45	122	0.009	0.002	0.001	Zimbabwe	11	0	0	0.001	0.001
Samoa	12	7	0	0	0.001						

Note: The table is arranged in descending PageRan centrality order. (*In-D: In-degree*; *Out-D: Out-degree*; *BC: Betweeness*; *EVC: Eigenvector*; *PR: PageRank*)

A.3 Classification of policy domains

To categorize policy documents thematically, we employed an automated annotation workflow using a large language model (LLM). We based our classification scheme on the Comparative Agendas Project (CAP) framework (Baumgartner, Breunig, and Grossman, 2019), assigning each document to one of the CAP's predefined policy domains. The classification relied on the summary metadata provided by Overton.

Following recent recommendations emphasizing transparency and reproducibility in computational research (Barrie, Palmer, and Spirling, 2025), we utilized Meta's open-source Llama3.3-70B model, hosted locally. This choice ensured greater control and auditability, while maintaining performance on par with commercial LLM offerings.

We used a consistent, structured prompt format to guide the model's outputs. Each input comprised two parts:

- System instruction: A fixed prompt that described the classification task and listed all 21 CAP categories along with brief definitions to reduce ambiguity.
- User content: The summary of the policy document to be classified.

Each summary was evaluated independently in a clean model session to eliminate carryover effects. The model was directed to respond with only the name of the most appropriate policy category, omitting any additional commentary.

An example of the instruction prompt used in all sessions is provided below:

> **System prompt:** Your task is to categorize a government document summary from {`Country name`} into one of the following policy topics. Return only the name of the single most-likely category. The categories are:
>
> - `Macroeconomics`: Domestic economic policy.
> - `Civil Rights`: Civil and minority rights.
> - `Health`: Healthcare policy and funding.
> - `Agriculture`: Farming and agricultural policy.
> - `Labor`: Employment, labor rights, and pensions.
> - `Education`: Education policy and funding.
> - `Environment`: Environmental regulations and policy.
> - `Energy`: Energy policy and regulation.
> - `Immigration`: Immigration, refugees, and citizenship.
> - `Transportation`: Transportation policy and infrastructure.
> - `Law and Crime`: Law enforcement, crime, and justice.
> - `Social Welfare`: Social assistance and welfare programs.
> - `Housing`: Housing and urban affairs.
> - `Domestic Commerce`: Business and trade within the country.
> - `Defense`: National defense and military policy.
> - `Technology`: Science, technology, and communication policy.
> - `Foreign Trade`: Trade agreements and international commerce.
> - `International Affairs`: Diplomacy, foreign aid, and global relations.
> - `Government Operations`: Public administration and agencies.
> - `Public Lands`: Land management, water policy, and territorial issues.
> - `Culture`: Arts, culture, and heritage policy.
>
> Return your answer in the following JSON format:
>
> {'policy_domain': 'Policy domain category from the options above'}
>
> Input: {`Summary of policy document`}

Each document underwent this classification protocol individually to maintain consistency and reproducibility across the dataset.

A.3.1 Validation Protocol

To validate the accuracy and reliability of our automated policy domain classification, we implemented a protocol involving both human annotation and a comparative assessment with an established external model. For this purpose, a random sample of 1,000 policy document summaries was drawn from our dataset.

Each of these 1,000 documents was manually annotated by a human expert. The annotator was instructed to identify the "most likely" policy domain based on the identical CAP framework and definitions used by the LLM, ensuring a consistent application of classification logic. This manual annotation served as our gold standard for evaluation.

Concurrently, these same 1,000 policy documents were also classified using a pretrained model `poltextlab/xlm-roberta-large-english-cap-v3` from the CAP Babel Machine (Sebők et al., 2024). This allowed for a direct, quantitative comparison of our Llama3.3-70B model's performance against an established encoder-based model specifically trained for CAP classification.

We then assessed the agreement between our Llama3.3-70B model's classifications and the human annotator's judgments. The LLM's derived categorizations coincided with the human annotator's in 78.9% of the instances. To provide a robust measure of inter-rater reliability, Cohen's Kappa was calculated, yielding (κ=0.768; z-value: 81.6; p-value: < 0.001). This result indicates substantial agreement between our automated classification and human judgment.

For direct comparison, we also calculated the agreement between the human annotator and the CAP Babel Machine model. Their classifications agreed in 68.8% of the cases. The corresponding Cohen's Kappa was (κ=0.660; z-value: 73.4; p-value: < 0.001).

These validation results underscore the strong performance of the Llama3.3-70B model and, notably, its superior accuracy when compared to the established CAP Babel Machine, further supporting its suitability for this large-scale classification task.

Table A5: Comparative Agendas Project (CAP) policy domains and descriptions

Policy Domain	Description
1. Macroeconomics	Issues related to general domestic macroeconomic policy.
2. Civil Rights	Issues related generally to civil rights and minority rights.
3. Health	Issues related generally to health care, including appropriations for general health care government agencies.
4. Agriculture	Issues related to general agriculture policy, including appropriations for general agriculture government agencies.
5. Labor	Issues generally related to labor, employment, and pensions, including appropriations for government agencies regulating labor policy.
6. Education	Issues related to general education policy, including appropriations for government agencies regulating education policy.
7. Environment	Issues related to general environmental policy, including appropriations for government agencies regulating environmental policy.
8. Energy	Issues generally related to energy policy, including appropriations for government agencies regulating energy policy.
9. Immigration	Issues related to immigration, refugees, and citizenship.
10. Transportation	Issues related generally to transportation, including appropriations for government agencies regulating transportation policy.
12. Law and Crime	Issues related to general law, crime, and family issues.
13. Social Welfare	Issues generally related to social welfare policy.
14. Housing	Issues related generally to housing and urban affairs.
15. Domestic Commerce	Issues generally related to domestic commerce, including appropriations for government agencies regulating domestic commerce.
16. Defense	Issues related generally to defense policy, and appropriations for agencies that oversee general defense policy.
17. Technology	Issues related to general space, science, technology, and communications.
18. Foreign Trade	Issues generally related to foreign trade and appropriations for government agencies generally regulating foreign trade.
19. International Affairs	Issues related to general international affairs and foreign aid, including appropriations for general government foreign affairs agencies.
20. Government Operations	Issues related to general government operations, including appropriations for multiple government agencies.
21. Public Lands	Issues related to general public lands, water management, and territorial issues.
23. Culture	Issues related to general cultural policy issues.

Table A6: Distribution of policy domains classes by region.

Policy domain	**Pooled**	AG	EEG	APG	GRULAC	WEOG	USA
Health	20.68% (257481)	37.51% (3,553)	24.57% (5,404)	23.97% (14,142)	11.78% (6,007)	22.85% (124,465)	18.59% (103,910)
Environment	17.47% (217550)	13.30% (1,260)	14.05% (3,089)	15.71% (9,266)	10.07% (5,133)	17.05% (92,841)	18.95% (105,961)
Government Operations	8.05% (100288)	5.98% (566)	6.58% (1,448)	2.48% (1,464)	16.73% (8,527)	4.95% (26,938)	10.97% (61,345)
Education	7.20% (89594)	7.52% (712)	13.85% (3,045)	10.76% (6,349)	10.60% (5,403)	6.99% (38,089)	6.44% (35,996)
Transportation	5.11% (63670)	1.40% (133)	1.36% (299)	2.00% (1,181)	1.22% (622)	7.16% (38,967)	4.02% (22,468)
Law and Crime	4.59% (57193)	1.82% (172)	3.16% (694)	1.95% (1,152)	3.28% (1,671)	3.36% (18,282)	6.30% (35,222)
Macroeconomics	4.51% (56107)	9.69% (918)	6.52% (1,433)	5.29% (3,123)	14.64% (7,466)	5.23% (28,480)	2.63% (14,687)
Labor	3.85% (47932)	1.34% (127)	2.51% (551)	2.55% (1,507)	3.53% (1,801)	3.67% (19,965)	4.29% (23,981)
Agriculture	3.35% (41682)	5.50% (521)	3.05% (671)	9.51% (5,609)	3.84% (1,960)	4.49% (24,452)	1.51% (8,469)
Housing	3.00% (37395)	1.31% (124)	1.08% (238)	0.78% (463)	0.74% (375)	2.12% (11,551)	4.41% (24,644)
Energy	2.98% (37083)	1.60% (152)	1.47% (323)	2.93% (1,729)	4.84% (2,465)	2.68% (14,587)	3.19% (17,827)
Domestic Commerce	2.93% (36516)	2.08% (197)	1.44% (316)	3.91% (2,307)	2.25% (1,146)	2.22% (12,084)	3.66% (20,466)
Social Welfare	2.81% (34957)	1.60% (152)	2.46% (542)	1.55% (912)	2.34% (1,191)	2.97% (16,171)	2.86% (15,989)
Civil Rights	2.10% (26182)	2.01% (190)	2.56% (562)	1.45% (855)	3.41% (1,739)	2.64% (14,374)	1.51% (8,462)
Technology	2.09% (26041)	2.31% (219)	4.01% (881)	5.93% (3,499)	2.44% (1,246)	2.09% (11,369)	1.58% (8,827)
No category	2.01% (25043)	0.87% (82)	1.15% (252)	1.88% (1,108)	1.12% (571)	2.25% (12,243)	1.93% (10,787)
Public Lands	1.78% (22141)	0.48% (45)	0.21% (47)	0.18% (109)	0.33% (170)	0.50% (2,745)	3.40% (19,025)
Culture	1.58% (19680)	1.11% (105)	1.52% (335)	3.10% (1,826)	2.56% (1,306)	1.89% (10,305)	1.04% (5,803)
International Affairs	1.43% (17795)	1.40% (133)	3.29% (723)	1.83% (1,079)	2.23% (1,135)	1.98% (10,787)	0.70% (3,938)
Immigration	0.94% (11765)	0.23% (22)	0.88% (193)	0.44% (260)	0.54% (275)	1.35% (7,362)	0.65% (3,653)
Defense	0.93% (11612)	0.10% (9)	3.62% (797)	0.45% (265)	0.21% (105)	0.78% (4,229)	1.11% (6,207)
Foreign Trade	0.59% (7385)	0.83% (79)	0.68% (149)	1.33% (783)	1.31% (668)	0.79% (4,321)	0.25% (1,385)

A.4 Variation across regions

Having established broad patterns about the distribution of policy-based and academic references in the main body, we provide a more disaggregated examination into how these dynamics vary across regions. Specifically, we use references in policy documents to analyze differences in citation practices between UN regional groups. These groupings reflect geopolitical affiliations and geographic proximity. To complement this perspective, we also distinguish between references that originate from 'developed' economies—typically high-income countries with robust research infrastructures—and those from other regions. This dual approach allows us to disentangle factors like spatial proximity and economic stratification relation to patterns of knowledge use.

On the policy side, we present results from models assessing the probability that a reference targets a foreign government, and whether such references are directed toward neighboring countries, countries within the same regional group, or developed' economies. On the academic side, we evaluate regional differences in the likelihood that a scholarly citation refers to a paper authored exclusively by researchers from developed' economy institutions, and whether at least one author is affiliated with an institution in the citing country.

Our findings reveal notable regional heterogeneity in policy citations and more modest variation in scholarly citations. For instance, while references to foreign government documents are roughly equally likely across most regions, African policy documents are significantly less likely—by about half—to cite foreign governments. Further, when it comes to citing neighboring countries, we observe no significant differences between Eastern European, Western European and Others, and Latin American and Caribbean countries. In contrast, Asia-Pacific and African countries are significantly less likely to cite their immediate neighbors. The starkest differences emerge in citations to governments within the same regional group: the likelihood of such references is 3.5 times greater in the Western European and Others group compared to Latin American and Caribbean countries. This is in contrast to the scholarly-end, where the most notable gap is a higher likelihood—approximately 11% greater—of citing research produced exclusively by 'developed' economy institutions in documents from the Western European and Others group.

Figure A3 provides an overview of these results. The estimates represent relative risks, capturing the ratio of the probability of observing a reference with a given characteristic in a UN regional group compared to the probability of observing it in documents authored by Latin American and Caribbean countries. While the geographic distribution of references highlights asymmetries in whose knowledge reaches the policy sphere, another key dimension of variation can emerge when examining differences across policy domains.

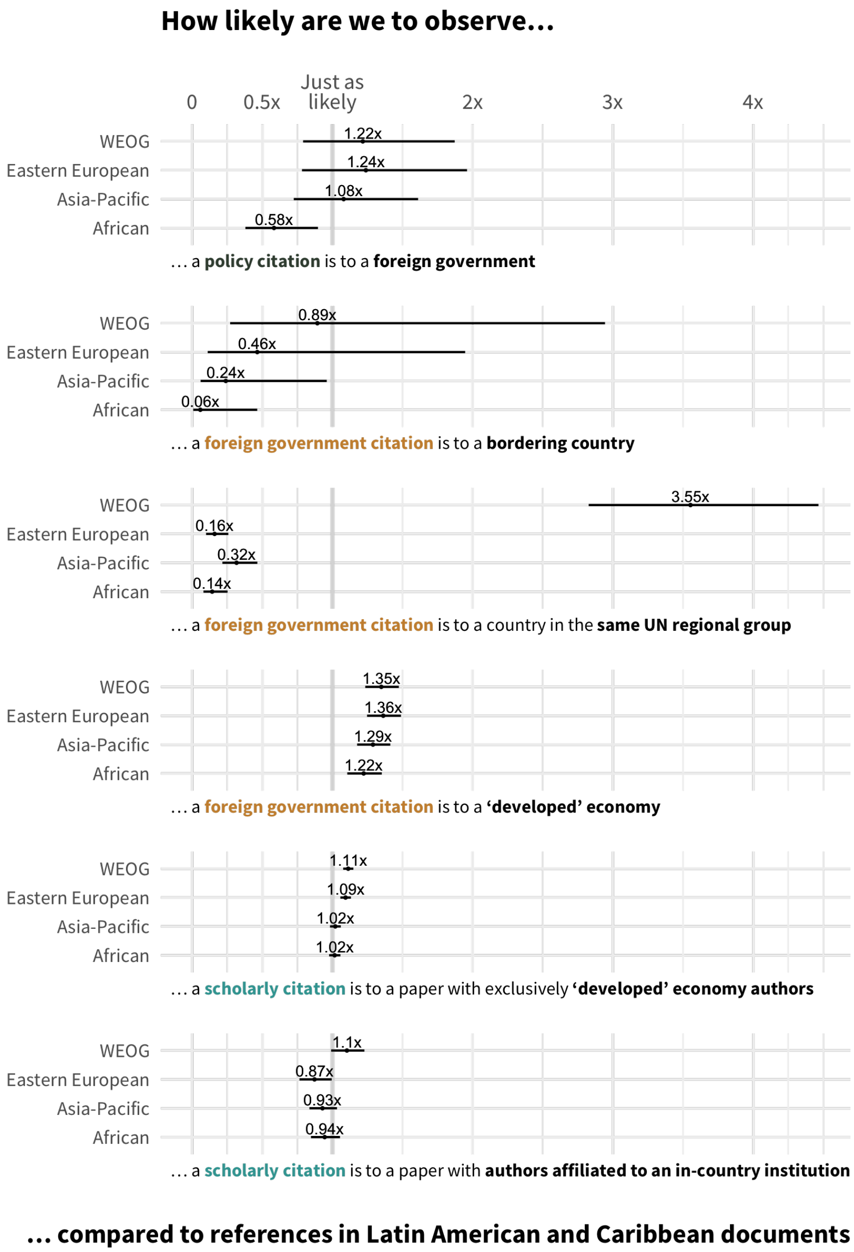


Figure A3: Marginal effects on observing a reference with a specified feature across groups. Results from six logistic mixed-effects models with document government author random effects. The estimates in the figure are relative risks representing the ratio of probability of observing a reference with the characteristic in a UN region group to the probability of observing it in documents authored by Latin American and Caribbean countries.

References

Arel-Bundock, Vincent, Nils Enevoldsen, and CJ Yetman (2018). "countrycode: An R package to convert country names and country codes". *Journal of Open Source Software* 3.28, p. 848.

Arel-Bundock, Vincent, Noah Greifer, and Andrew Heiss (2024). "How to Interpret Statistical Models Using marginaleffects for R and Python". *Journal of Statistical Software* 111.9, pp. 1–32.

Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, Toby Hocking, and Benjamin Schwendinger (2024). *data.table: Extension of 'data.frame'*.

Barrie, Christopher, Alexis Palmer, and Arthur Spirling (2025). *Replication for Language Models: Problems, Principles, and Best Practice for Political Science*.

Baumgartner, Frank R., Christian Breunig, and Emiliano Grossman, eds. (2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press.

Brooks, Mollie E., Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler, and Benjamin M. Bolker (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling". *The R Journal* 9.2, pp. 378–400.

Csárdi, Gábor, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller (2025). *igraph: Network Analysis and Visualization in R*.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton (2019). *xtable: Export Tables to LaTeX or HTML*.

Firke, Sam (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy (2024). *gt: Easily Create Presentation-Ready Display Tables*.

Lin, Hause and Tawab Safi (2024). "ollamar: An R package for running large language models". *PsyArXiv*.

Massicotte, Philippe and Andy South (2023). *rnaturalearth: World Map Data from Natural Earth*.

Massimo, Aria, Le Trang, Cuccurullo Corrado, Belfiore Alessandra, and Choe June (2024). "openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex". *The R Journal* 15 (4), pp. 167–180.

Nenadic, O. and M. Greenacre (2007). "Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package". *Journal of Statistical Software* 20.3, pp. 1–13.

Ooms, Jeroen (2014). "The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects". *arXiv:1403.2805 [stat.CO]*.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Sebők, Miklós, Ákos Máté, Orsolya Ring, Viktor Kovács, and Richárd Lehoczki (2024). "Leveraging Open Large Language Models for Multilingual Policy Topic Classification: The Babel Machine Approach". *Social Science Computer Review*, p. 08944393241259434.

South, Andy, Schramm Michael, and Philippe Massicotte (2024). *rnaturalearthdata: World Vector Map Data from Natural Earth Used in 'rnaturalearth'*.

van den Brand, Teun (2024). *ggh4x: Hacks for 'ggplot2'*.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

— (2023). *httr: Tools for Working with URLs and HTTP*.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan (2023). *dplyr: A Grammar of Data Manipulation*.

Wickham, Hadley and Lionel Henry (2025). *purrr: Functional Programming Tools*.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich (2024). *tidyr: Tidy Messy Data*.

Worldwide Governance Indicators, 2024 Update, (2025). World Bank, Accessed on 2025-02-18.

The Bundestag Expert Witness Tracker (BEWIT): A database of German Bundestag public expert hearings

Sebastian Ramirez-Ruiz (Hertie School)*

Abstract. In democratic systems, legislators tackle complex policy challenges while juggling limited time, attention, and information, alongside pressures from constituents and interest groups. Although they frequently rely on external expertise, we still know surprisingly little about who is formally invited to contribute knowledge to the legislative process. This paper introduces the Bundestag Expert Witness Tracker (BEWIT), a novel, hand-curated dataset documenting over 11,000 expert-affiliation pairs from 1,804 public committee hearings in the German Bundestag between 2009 and 2024. The database captures detailed metadata on hearings, expert identities, organizational affiliations, and—where available—linked *Lobbyregister* records and academic researcher profiles. These hearings offer a rare, observable arena to study how legislators curate informational inputs across diverse policy domains. To demonstrate the empirical potential of BEWIT, I present two applications: one that analyzes the disciplinary and institutional composition of academic experts in committee hearings, and another that examines whether a procedural reform in the middle of the 20th legislative period—requiring disclosure of which party invited each expert—correlates with shifts in the makeup of the witness pool. This resource opens new avenues for investigating how democratic institutions filter expertise, navigate competing demands, and structure the informational foundations of policymaking.

Keywords. Public committee hearings; Legislative data; German Bundestag; Evidence-based policymaking; Expert-information; Interest groups; Knowledge utilization

*Corresponding author. Address: Friedrichstrasse 180, 10117 Berlin, Germany. Email: ramirez-ruiz@hertie-school.org, Web: seramirezruiz.github.io.

Information plays a pivotal role in democratic decision-making. It can influence how elected officials understand policy problems, the range of options considered, and which solutions gain support. In that sense, the sources legislators rely on for information can play a decisive role in shaping policy outcomes (Dunlop, 2016). Despite the vast attention paid to the general role of information in democracies (e.g., Carpini and Keeter, 1993, 1996; Broockman and Skovron, 2018; Jerit and Zhao, 2020), our understanding of how information and expertise permeate legislative practice through legislators' interactions with experts and external groups remains limited.

Although research at the intersection of public decision-making, evidence, and expertise has gained salience in recent decades, it has disproportionately focused on the executive branch (see Oliver et al., 2014). Yet legislators face a distinct set of challenges in navigating expert input. They have to balance strategic, symbolic, and substantive uses of information in a highly visible political arena. Within legislative settings, scholars have often examined how information is disseminated by specific actors—such as interest groups, bureaucrats, or researchers—permeating policy, typically treating them in isolation and overlooking the interactions between them. Existing studies have identified institutional- and individual-level factors shaping how academic research evidence enters the legislative process (Ouimet et al., 2023, for an overview). Still, much of this work underplays the agency legislators have in selecting their sources. As active curators of expertise (Walgrave and Dejaeghere, 2017), legislators make deliberate choices about where to turn for information and whose voices to prioritize.

This study introduces a new database to help address that gap and contribute to a growing body of research on how legislatures are informed and who legislators turn to for information (e.g., Halpin, MacLeod, and McLaverty, 2012; Geddes, 2018; Ban, Park, and You, 2023; Vallejo Vera, 2023). The data present a comprehensive record of 1,804 public committee hearings with experts in the German Bundestag between 2009 and 2024, comprising over 10,000 expert participants. These hearings—*Öffentliche Sachverständigenanhörungen*—are convened when bills are complex, contested, or politically salient, and offer a unique, observable arena to examine how legislators interact with external expertise. Unlike informal or internal consultations, these hearings leave a public record of which voices are invited to speak and, more recently, which party group extended the invitation.

The data set enables scholars to ask questions about the flow of expertise in legislative contexts: Who gets invited to testify? What kinds of knowledge are seen as authoritative across policy areas? How do institutional changes in transparency affect the composition of expert input? I illustrate the utility of the data by exploring two applications. First, I describe the overall composition of the expert pool with attention to the presence of academic researchers relative to other types of actors, examining variation in disciplinary background and committee context. Second, I leverage a procedural change in the 20th legislative session that made the source of each invitation public to explore whether the transparency change is associated with shifts in the patterns of expert selection.

Together, these data offer a new lens for studying how legislators curate information, how expertise is platformed in democratic institutions, and how procedural rules shape the informational foundations of lawmaking. Given the challenges to studying the preferences for, use, and sharing of expertise in legislatures due to the unobservability and secretive nature of legislative practice, public committee hearings provide a valuable empirical window into how elected officials navigate the tradeoffs between expertise, partisanship, and legitimacy in the policymaking process.

Information, political elites, and legislative practice

Information plays a central role in the policymaking process, not only as a raw input for decision-making but also as a structuring force that shapes how problems are perceived, what solutions are considered viable, and which actors are deemed legitimate contributors. From theories of bounded rationality to models of policy learning, agenda-setting, and use of evidence, information is treated as both a constraint and an enabler of political behavior (e.g., Simon, 1944; Lindblom, 1959; Weiss, 1979; Jones and Baumgartner, 2005).

For political elites, especially legislators in democratic systems, navigating the information environment presents a unique set of challenges. Their job requires them to legislate across an increasingly complex policy landscape while contending with limited time, cognitive bandwidth, and the need to ensure the political acceptability of their decisions (Webber, 1987). Generally, legislators are not themselves subject-matter experts, yet they are expected to craft laws in areas that often demand specialized knowledge. As a result, they must rely on others—bureaucrats, interest groups, academics, and various stakeholders—to provide information and guidance, often within highly politicized and uncertain contexts (De Bruycker, 2016; Cross et al., 2021; Bellodi, 2023; Ban, Park, and You, 2025).

A conceivable consequence of this balancing act between legislative constraints and the pressure to make informed decisions is the risk of informational overload (Jones and Baumgartner, 2005; Senninger and Seeberg, 2024). Evidence from studies of party leaders suggests that legislators do, in fact, experience such overload and have developed strategies to manage it. These include both passive and active filtering—outsourcing information selection to institutional procedures or legislative staff, as well as actively curating sources to align with political or policy objectives (Walgrave and Dejaeghere, 2017). If information is so central to legislative decision-making and if legislators must filter what they consume, it becomes essential to ask: whose information do they select? And, whose voices are given a platform in the legislative process?

Scholars have identified several pathways through which information enters the legislative arena, each rooted in different theoretical traditions. One prominent avenue is policy learning, where policymakers seek out information to update their beliefs or approaches in response to uncertainty, new evidence, or perceived policy successes and failures (Sanderson, 2002; Meseguer, 2005; Dunlop and Radaelli, 2018; Pereira, 2022). In these instances, information can serve a corrective function, helping legislators make more effective or technically sound decisions.

Another well-established stream of literature deals with informational lobbying, where interest groups and other organized actors provide targeted knowledge to legislators in an effort to influence outcomes (Austen-Smith, 1993; Contandriopoulos et al., 2010; Klüver, 2012; Chalmers, 2013; De Bruycker, 2016; Schnakenberg and Turner, 2019; Albareda, 2020). In this case, expertise is a strategic resource exchanged in the policy process—valuable not only for its content, but for its ability to frame issues and shape legislative agendas.

Legislators also rely on constituent preferences as a form of informational input. Signals from voters, often mediated through polling, public feedback, or media coverage, can strongly influence elite decision-making, particularly on salient or controversial topics (Butler and Nickerson, 2011; Broockman and Skovron, 2018; Chu and Recchia, 2022). In such cases, the informational role of public opinion is less about technical accuracy and more about political viability.

Further, political elites often engage in proactive information seeking, particularly in policy domains that are highly technical or complex. The likelihood of such behavior can vary by institutional context, policy domain, and the legislator's own level of interest or expertise (Landry, Lamari, and Amara, 2003; Loewen, Rubenson, and McAndrews, 2022; Ramirez-Ruiz, 2025). Experimental research has shown that policy professionals are far from being merely passive recipients of expertise—they are active curators of information, influenced by both strategic considerations and personal motivations (Baekgaard et al., 2019; Pereira and Öhberg, 2020; Lee, 2022; Senninger and Seeberg, 2024; Vivalt, Coville, and Kc, 2025).

To further understand how these dynamics play out in practice, we must also consider the timing and source of information in relation to the legislative process itself. Recent research suggests that some types of information are more salient at different stages of the policy-making process. Public opinion and media signals tend to shape issue prioritization during the agenda-setting phase, whereas interest group input and expert knowledge become more influential during policy formulation (Willems, Maes, and Walgrave, 2024). In these later stages, legislators often rely on expert information not only for its substantive value but also for its symbolic utility—drawing on specialized knowledge to bolster the credibility of their proposals and signal competence to colleagues and constituents (Fagan and McGee, 2022; Hünermund, 2024).

The tensions between the strategic, symbolic, and substantive uses of expert information by political elites are often difficult to observe directly. However, we can gain some insight by looking at the organizational settings in which these interactions take place. Legislative committees, in particular, play a crucial role in gathering, evaluating, and legitimising information during the policymaking process. As key sites of policy formulation, they not only channel expertise but also make visible the processes through which certain forms of information are selected over others. Scholars have long emphasized the informational function of committees (Krehbiel, 1992; Baron, 2000; Hirsch and Shotts, 2012; Ban, Park, and You, 2024), highlighting their ability to reduce collective decision-making costs by organizing knowledge along specialized lines. Through expert invitations, public hearings, and the production of reports, committees help determine not just what kinds of information enter the legislative process, but also whose voices are granted authority within it.

If legislatures are arenas for the negotiation of knowledge and power, then committee hearings offer a rare, observable window into how that negotiation is staged. To make these dynamics empirically traceable, I turn to a concrete case: public committee hearings with experts in the German Bundestag. These hearings represent a key institutional site where legislators engage external expertise, make informational choices visible, and shape the direction of legislative decisions through the selection of voices granted a formal platform.

Public committee hearings with experts in the Bundestag

Public committee hearings represent a central institutional mechanism through which legislators engage with external expertise. Across democratic systems—including the U.S. Congress, the UK Parliament, and the German Bundestag—committees convene hundreds of such hearings annually. The study of committee hearings holds a longstanding tradition within the broader field of legislative studies (e.g., Diermeier and Feddersen, 2000). Scholars have used hearing data to investigate questions ranging from agenda-setting, interest group access, and the dynamics of political polarization.

All-in-all, public hearings serve a crucial function in supporting legislative information acquisition, a foundational task for committees (Krehbiel, 1992; Strøm, 1998; Baron, 2000). While procedural formats might be different across countries, the primary goal of such hearings tends to be informational: committees elicit expert testimony to inform the legislative process. Notably, the German Bundestag's Rules of Procedure underscore this function explicitly:

> For the purpose of obtaining information on a subject under debate, a committee may hold public hearings of experts, representatives of interest groups and other persons who can furnish information (Deutscher Bundestag, 2022).

Unlike the U.S. Congress, where hearings often serve additional purposes such as oversight or confirmations (Ban, Park, and You, 2023), Bundestag hearings are more narrowly oriented toward issue-specific deliberation and policy formulation.

That said, the informational logic of hearings might not be the only one at play in the Bundestag. Hearings also serve strategic and symbolic functions. Previous evidence from the U.S. context suggests that committee hearings can also serve as stages for political grandstanding, where legislators use the public forum to signal positions to voters, media, or party leaders (Park, 2017, 2023). Specific to the Bundestag, Hünermund (2024) suggests that these hearings also serve as signals to democratic legitimacy by visibly engaging outside expertise and offering a venue for pluralistic input. In this sense, public hearings serve a dual role: they are both tools for gathering information and instruments for signalling accountability, responsiveness, or partisan alignment.

There is valuable scholarship on how legislative hearings function in the German context. For instance, Geddes (2024) offers deep qualitative insights by comparing knowledge practices across the UK and German parliaments, highlighting institutional differences in how expertise is gathered and evaluated. Complementing this, descriptive research shows that associations and organized interests constitute a significant share of witnesses in Bundestag committee hearings (Dhungel and Linhart, 2014). Yet, despite longstanding theoretical interest in the role of information in legislatures, we still know relatively little about the actual flow of expert input beyond these organized actors. Who is invited from this broader group, what perspectives they bring, and how inclusive or selective these fora are remain open empirical questions.

Crucially, witness selection in the Bundestag is not a neutral process. It reflects the technical needs of committees, party preferences, strategic considerations, and broader institutional norms. Unlike in the UK Parliament—where committee staff play a central role in identifying and inviting witnesses—the Bundestag delegates this responsibility to the parliamentary party groups (Geddes, 2024). Each group nominates a share of witnesses proportional to its parliamentary strength. As a result, parties must actively navigate the tension between seeking relevant expertise and advancing their political priorities.

That said, public committee hearings offer a valuable and observable arena for examining how legislators interact with external sources of information. They shed light on how elected officials navigate the tension between seeking knowledge, advancing political goals, and maintaining legitimacy. When examining patterns in witness invitations in the Bundestag, we can gain insight into which forms of expertise are elevated, which voices are granted a platform, and how informational access is structured in parliamentary practice. This study contributes to that effort by introducing a new dataset that systematically documents expert testimony in the German Bundestag.

The Bundestag Expert Witness Tracker (BEWIT)

Data collection. I constructed a new database documenting all public expert hearings held by committees of the German Bundestag across four legislative sessions, from 2009 to 2024. The database is built on open information hosted by the Bundestag's official web archive, which provides structured records of legislative activity for each session. I collected the full list of committee seatings across the 17th to 20th legislative periods (2009–2024) by scraping the archive. I extracted all records in which the official documentation lists the agenda items as *Öffentliche Anhörung* (public hearing) as part of the protocol. In total, the database includes 1,804 unique hearings across the 25 standing committees of the Bundestag. For each hearing, I collected metadata such as the hearing title, date, legislative files under discussion, the initiating party, and—when available—the full transcript provided by the Bundestag's documentation office. Figure 1 presents the distribution of hearings across committees and over time.

A central output rendered by the Bundestag's documentation office in the context of public committee hearings with experts and interest groups is a list of experts. This list is a transparency. While occasional mismatches or classification errors cannot be entirely ruled out, the formalized nature of the source and the double-coding process substantially reduce the risk of systematic error.

External data integration. To further enrich the database and broaden its utility for empirical research, I linked expert-level records to two auxiliary data sources. This linking effort allows for more granular analyses of the types of expertise legislators draw on, as well as the characteristics of the actors supplying that expertise.

For academic participants, I connect the entries for individual experts to researcher profiles in OpenAlex (Priem, Piwowar, and Orr, 2022), a comprehensive, open-source index of scholarly work. To link academic expert witnesses to their scholarly profiles, I extracted all expert-affiliation pairs classified under a research institution and used the resulting name strings to query the OpenAlex "Authors" API endpoint. The returned records were merged with the original affiliation data from the Bundestag hearings. I initially screened for one-to-one matches based on both full names and current institutional affiliations. In cases where the most recent institutional affiliation did not match the hearing record, I examined past affiliations listed in OpenAlex to identify likely matches. When ambiguity remained, I conducted manual web searches to verify identities—either by locating institutional profiles or by searching for publication titles that could be used to query the OpenAlex "Works" API. This process allowed me to resolve cases where name variants or outdated affiliations might otherwise obscure matches.

After this multi-step validation and linkage procedure, 1,879 out of 2,085 academic expert-affiliation pairs (approximately 90%) were successfully linked to a unique OpenAlex author identifier. These correspond to 1,701 unique OpenAlex IDs, as some researchers were invited under different affiliations or appeared multiple times over the covered period. Each linked researcher profile enriches the BEWIT database with structured metadata on disciplinary background, publication record, and citation profile, enabling more nuanced investigations of academic engagement in legislative contexts.

These metadata open up new possibilities for investigating how scholarly expertise is represented across policy domains—for example, whether committee invitations skew toward certain disciplines, how research salience or academic rank relates to legislative engagement, or how epistemic authority is institutionally distributed.

For representatives of interest groups and private corporations, I connect their organizational affiliations to the Bundestag's official *Lobbyregister*, a public registry that tracks the lobbying ecosystem in Germany. The register includes detailed information on each organization's described aims, financial structure, funding sources (including public grants), and registered representatives. By connecting the expert-level data to this registry, the database facilitates a more transparent view of the organizational and resource capacities behind expert testimony allowing, for instance, analyses of how well-resourced or broadly represented groups influence committee discussions across domains.

To accomplish this links, I started with all affiliations categorized as "Interest groups and nonprofits" and "Private sector" in the database. Using the names of these organizations, I conducted manual searches on the Lobbyregister's official web portal to identify corresponding entries. For each match, I recorded the organization's unique Lobbyregister ID into a separate data table. I then performed a double-check procedure to verify that each ID correctly corresponded to the intended organization.

This process yielded 1,261 matched IDs for interest groups and nonprofits, covering approximately 50% of such affiliations in the database, and 221 matched IDs for private corporations, accounting for roughly 30% of the relevant entries. The Lobbyregister does not comprehensively cover all relevant actors: under current legislation, registration is only mandatory for representatives of special interests who meet specific legal criteria, and many actors—such as those operating episodically, under exemptions, or below reporting thresholds—may not be registered as it would be voluntary. This partially limits coverage, but matched cases skew toward the most visible and resourceful actors, which are often those of greatest interest for analyzing lobbying capacity and legislative influence. In total, 55% of expert appearances—6,377 of 11,585—from members of these two groups have a linked Lobbyregister ID.

This integration enables researchers to trace the organizational characteristics and resource bases of actors providing expert testimony, facilitating investigations into how well-resourced or broadly represented groups engage with legislative committees.

To evaluate the possibility of systematic missingness in the linkage process, I estimated logistic regression models predicting the likelihood of a successful match to an OpenAlex or Lobbyregister identifier based on the year of the parliamentary hearing (see Supplementary Table B2).

For academic participants, there was no consistent or strong evidence of time-based bias in OpenAlex matches. While match rates were significantly higher in 2020 and 2022 compared to 2010 (OR = 2.30, 95% CI = [1.05, 4.86], p = 0.031), (OR = 2.24, 95% CI = [0.99, 4.94], p = 0.047), these increases were not observed uniformly across years.

For interest group and private sector participants, no year showed a statistically significant difference in Lobbyregister match rates, suggesting that linkage completeness was stable over time and not influenced by legislative period. These findings indicate that the linkage process does not appear to be driven by the recency of the meetings.

Together, these auxiliary connections substantially extend the analytical scope of the database. They not only allow researchers to disaggregate expert testimony by category but also to explore the reputational, disciplinary, and financial dimensions of actors that shape the informational environment of legislative decision-making.

Database structure and access. The Bundestag Expert Witness Tracker (BEWIT) is implemented as a relational database, enabling modular and scalable analysis. Figure 2 illustrates the database schema. The core table, *list of experts*, is connected to three auxiliary tables: *hearings* (containing hearing-level metadata and permanent links to the verbatim meeting protocols), *academics* (with linked OpenAlex researcher profiles), and *affiliations* (providing institutional data for interest groups, government bodies, and other organizations). Unique identifiers serve as keys to enable consistent joining across tables.

This relational structure not only facilitates flexible queries and disaggregation across different dimensions of the data but also makes the database more robust to future updates and extensions—for instance, by adding new legislative sessions, enriching actor metadata, or incorporating additional sources. Similar to other efforts in the study of elite behavior that rely on relational data architectures to enable cumulative research and version control (e.g., Göbel and Munzert, 2022), the BEWIT is designed to support both immediate analysis and long-term expansion.

By combining high-quality official sources with manual verification and auxiliary linkage, BEWIT provides a transparent and extensible foundation for empirical research on legislative information flows. While it necessarily excludes informal consultations and closed-door processes, it offers rare visibility into the formal channels through which external expertise enters democratic policymaking.

A static version of the BEWIT database is publicly hosted on the Harvard Dataverse. The data is available in three formats widely used in the social sciences: comma-separated values (.csv), R binary files (.rds), and SQLite (.sqlite). In addition to access to standalone files, users will be able to interact with the database programmatically through the open-source `{bundestagexperts}` R package (*under construction*). The package will enable targeted access to the full database and is designed to integrate seamlessly into typical workflows in R.

Application 1: Characterizing academic experts present in public hearings

In the following sections, I demonstrate the utility of the database through two illustrative applications. These examples highlight the types of questions the database can help answer and provide an empirical overview of its content. The first application focuses on the academic experts who participate in public hearings, examining what disciplines they represent and how their presence varies across committees.

Interest in the use of research evidence in legislatures has grown in recent years (Ouimet et al., 2023), and the German Bundestag is no exception. While scholars have increasingly sought to assess the role of scientific input in parliamentary work, much of the existing evidence relies on self-reports from legislators and staff (Seidel et al., 2021; Riphahn and Schnitzer, 2022). The Bundestag Expert Witness Tracker (BEWIT) offers a complementary, observational perspective by systematically documenting when, where, and how researchers are invited to participate in committee deliberations. This kind of data is especially valuable in light of recent experimental evidence suggesting that direct interactions between legislators and academics can measurably increase the uptake of research in policymaking (Crowley et al., 2021). Understanding who is invited and under what conditions thus provides crucial insight into the informational foundations of legislative decision-making.

Figure 3 summarizes key patterns in the presence and profile of academic experts. Panel A shows their frequency relative to other groups, Panels B and D break down their disciplinary backgrounds, and Panel C shows variation across legislative committees. Researchers affiliated with universities, research institutes, and think tanks are the second most common group in the witness pool, only surpassed by representatives of interest groups and nonprofits.

Researchers are the second largest group invited to participate in Bundestag public committee hearings. Between 2009 and 2024, researchers accounted for approximately 20% of over 18,000 expert appearances in these committee hearings. This differs substantially from the U.S. context, where oversight and confirmations are frequent reasons hearings are convened. In the U.S., bureaucrats dominate the expert pool, making up over a quarter of

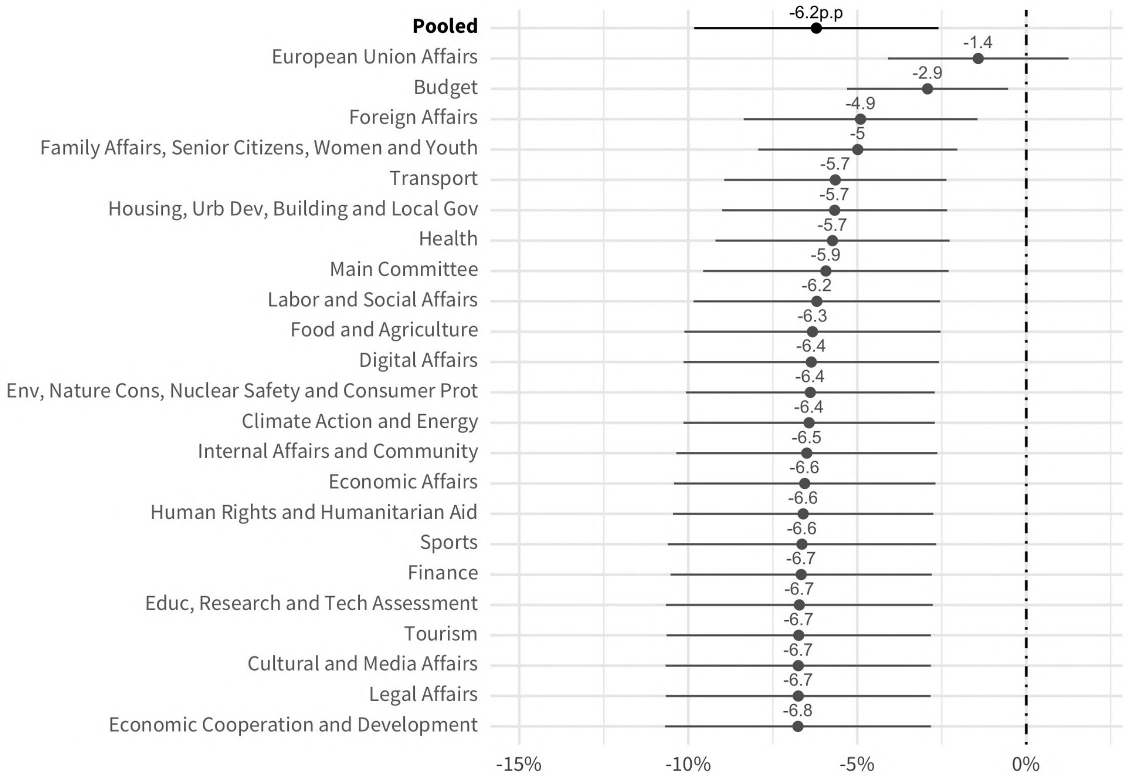


Figure 5: Predicted probability changes of an expert representing an interest group or nonprofit associated with the protocol change. Results from a logistic regression with committee fixed effects. The estimates in the figure report the average partial derivatives of the predicted probability along 95% confidence intervals. *N. obs* = 3,594.

Discussion

This paper introduced the Bundestag Expert Witness Tracker (BEWIT), a new database documenting over 18,000 expert appearances across more than 1,800 public hearings in the German Bundestag between 2009 and 2024. The BEWIT offers a novel resource for analyzing how legislatures access and organize external expertise in the policy process by systematically collecting and enriching data on expert testimony.

Through two illustrative applications, the paper demonstrated the analytical potential of this data source. First, it showed that academic researchers represent a substantial portion of expert witnesses—particularly in certain policy domains—offering empirical insight into how scientific expertise enters legislative deliberation. Second, it employed a recent procedural change to explore whether the availability of information around party-sponsored invitations is related to expert selection. The results suggest that increased transparency is associated with a modest decline in the presence of interest group representatives, pointing to responsiveness in party decisions.

Taken together, these findings underscore the complex and contested nature of expertise in parliamentary settings. Expert testimony is not merely an informational input but also a site of symbolic and strategic politics. Who is invited to speak reflects underlying institutional norms, partisan priorities, and, more generally, broader societal power structures.

The BEWIT opens several promising avenues for future research for scholars interested in the role of information, lobbying, and legitimacy in representative democracies. Legislative studies scholars, for example, can use the data to ask how expertise is distributed across different types of bills or policy domains, or whether party-specific witness selections reflect broader trends in parliamentary polarization. Scholars of evidence and expertise can explore how often independent academics are called upon, how their presence compares to organized interests, or how expert credentials vary over time or across issues. For researchers focused on interest groups and money in politics, the linked Lobbyregister data allows inquiries into the presence and resource profiles of corporate or nonprofit actors, offering insight into which types of groups gain informational access during policy formation. Beyond the Bundestag, the modular design of the database and the accompanying R package provide a template for comparative data collection in other legislatures.

Of course, the database also has some limitations. While the Bundestag has made significant efforts to document expert hearings, some historical records may be incomplete, and institutional affiliations are not always consistently standardized. Moreover, although the BEWIT captures who is invited to testify, it does not (yet) systematically record what is said during the hearings—or how it is said. The database includes permanent links to verbatim protocols in PDF format, where available, as well as to meeting landing pages that host video recordings for more recent legislative periods. Researchers are encouraged to build on the dataset by extracting text, audio, and video from these sources. In principle, such multimodal data—linked to identifiable entities such as MPs and experts—can enable richer analyses. For example, scholars could explore how different types of witnesses contribute substantively to deliberation, whether certain features of expert witnesses shape how they are addressed, track changes in discourse over time, or investigate whether MPs interact differently with in-group versus out-group experts based on party alignment, expertise type, or institutional background.

Ultimately, this paper contributes to a growing literature that seeks to open the "black box" of legislator relations to expert information. By making expert selection more empirically tractable, the BEWIT helps illuminate the informational foundations of democratic decision-making, and raises important questions about who speaks for knowledge in politics.

Acknowledgments

The author acknowledges financial support by the Hertie School in the form of a doctoral stipend. The author received no specific funding for this work.

References

Simon, Herbert A. (1944). “Decision-Making and Administrative Organization”. *Public Administration Review* 4.1, p. 16.
Lindblom, Charles E. (1959). “The Science of "Muddling Through"”. *Public Administration Review* 19.2, pp. 79–88.
Weiss, Carol H. (1979). “The Many Meanings of Research Utilization”. *Public Administration Review* 39.5, pp. 426–431.
Webber, David J. (1987). “Legislators’ Use of Policy Information”. *American Behavioral Scientist* 30.6, pp. 612–631.
Krehbiel, Keith (1992). *Information and Legislative Organization*. University of Michigan Press.
Austen-Smith, David (1993). “Information and Influence: Lobbying for Agendas and Votes”. *American Journal of Political Science* 37.3, pp. 799–833.
Carpini, Michael X. Delli and Scott Keeter (1993). “Measuring Political Knowledge: Putting First Things First”. *American Journal of Political Science* 37.4, pp. 1179–1206.
— (1996). *What Americans Know about Politics and why it Matters*. Yale University Press.
Strøm, Kaare (1998). “Parliamentary committees in european democracies”. *The Journal of Legislative Studies* 4.1, pp. 21–59.
Baron, David P. (2000). “Legislative Organization with Informational Committees”. *American Journal of Political Science* 44.3, pp. 485–505.
Diermeier, Daniel and Timothy J. Feddersen (2000). “Information and Congressional Hearings”. *American Journal of Political Science* 44.1, pp. 51–65.
Sanderson, Ian (2002). “Evaluation, Policy Learning and Evidence-Based Policy Making”. *Public Administration* 80.1, pp. 1–22.
Landry, Réjean, Moktar Lamari, and Nabil Amara (2003). “The Extent and Determinants of the Utilization of University Research in Government Agencies”. *Public Administration Review* 63.2, pp. 192–205.
Jones, Bryan D. and Frank R. Baumgartner (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.
Meseguer, Covadonga (2005). “Policy Learning, Policy Diffusion, and the Making of a New Order”. *The ANNALS of the American Academy of Political and Social Science* 598.1, pp. 67–82.
Contandriopoulos, Damien, Marc Lemire, Jean-Louis Denis, and Émile Tremblay (2010). “Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature”. *The Milbank Quarterly* 88.4, pp. 444–483.
Butler, Daniel M. and David W. Nickerson (2011). “Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment”. *Quarterly Journal of Political Science* 6.1, pp. 55–83.
Halpin, Darren, Iain MacLeod, and Peter McLaverty (2012). “Committee Hearings of the Scottish Parliament: Evidence Giving and Policy Learning”. *The Journal of Legislative Studies* 18.1, pp. 1–20.
Hirsch, Alexander V. and Kenneth W. Shotts (2012). “Policy-Specific Information and Informal Agenda Power”. *American Journal of Political Science* 56.1, pp. 67–83.
Klüver, Heike (2012). “Informational Lobbying in the European Union: The Effect of Organisational Characteristics”. *West European Politics* 35.3, pp. 491–510.
Chalmers, Adam William (2013). “Trading information for access: informational lobbying strategies and interest group access to the European Union”. *Journal of European Public Policy* 20.1, pp. 39–58.
Dhungel, Anna-Katharina and Eric Linhart (2014). “Interessenvermittlung in den Ausschüssen des Deutschen Bundestages”. *ZParl Zeitschrift für Parlamentsfragen* 45.4, pp. 743–762.
Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas (2014). “A systematic review of barriers to and facilitators of the use of evidence by policymakers”. *BMC Health Services Research* 14.1, p. 2.

De Bruycker, Iskander (2016). “Pressure and Expertise: Explaining the Information Supply of Interest Groups in EU Legislative Lobbying”. *JCMS: Journal of Common Market Studies* 54.3, pp. 599–616.

Dunlop, Claire A. (2016). “Knowledge, epistemic communities, and agenda setting”. In: *Handbook of Public Policy Agenda Setting*. Ed. by Nikolaos Zahariadis. Edward Elgar Publishing.

Green, Manda (2016). “The Parliamentary Women’s Delegation in France: Making Women’s Voices Heard?” *Parliamentary Affairs* 69.4, pp. 860–875.

Park, Ju Yeon (2017). “A Lab Experiment on Committee Hearings: Preferences, Power, and a Quest for Information”. *Legislative Studies Quarterly* 42.1, pp. 3–31.

Walgrave, Stefaan and Yves Dejaeghere (2017). “Surviving Information Overload: How Elite Politicians Select Information”. *Governance* 30.2, pp. 229–244.

Broockman, David E. and Christopher Skovron (2018). “Bias in Perceptions of Public Opinion among Political Elites”. *American Political Science Review* 112.3, pp. 542–563.

Dunlop, Claire A. and Claudio M. Radaelli (2018). “The lessons of policy learning: types, triggers, hindrances and pathologies”. *Policy & Politics* 46.2, pp. 255–272.

Geddes, Marc (2018). “Committee Hearings of the UK Parliament: Who gives Evidence and does this Matter?” *Parliamentary Affairs* 71.2, pp. 283–304.

Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen (2019). “The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians”. *British Journal of Political Science* 49.3, pp. 1117–1140.

Schnakenberg, Keith E. and Ian R. Turner (2019). “Signaling with Reform: How the Threat of Corruption Prevents Informed Policy-making”. *American Political Science Review* 113.3, pp. 762–777.

Albareda, Adrià (2020). “Prioritizing professionals? How the democratic and professionalized nature of interest groups shapes their degree of access to EU officials”. *European Political Science Review* 12.4, pp. 485–501.

Jerit, Jennifer and Yangzi Zhao (2020). “Political Misinformation”. *Annual Review of Political Science* 23.1, pp. 77–94.

Pereira, Miguel M. and Patrik Öhberg (2020). *The Expertise Paradox: How Policy Expertise Can Hinder Responsiveness*. preprint. Open Science Framework.

Cross, James P., Rainer Eising, Henrik Hermansson, and Florian Spohr (2021). “Business interests, public interests, and experts in parliamentary committees: their impact on legislative amendments in the German Bundestag”. *West European Politics* 44.2, pp. 354–377.

Crowley, D. Max, J. Taylor Scott, Elizabeth C. Long, Lawrie Green, Azaliah Israel, Lauren Supplee, Elizabeth Jordan, Kathryn Oliver, Shannon Guillot-Wright, Brittany Gay, Rachel Storace, Naomi Torres-Mackie, Yolanda Murphy, Sandra Donnay, Jenna Reardanz, Rebecca Smith, Kristina McGuire, Elizabeth Baker, Ana Antonopoulos, Mary McCauley, and Cagla Giray (2021). “Lawmakers’ use of scientific evidence can be improved”. *Proceedings of the National Academy of Sciences* 118.9, e2012955118.

Seidel, Katja, Hans Verbeek, Sophie Fessel, and Florian Meer (2021). *Nutzen von wissenschaftlicher Evidenz: Erwartungen an wissenschaftliche Expertise : Bericht über eine Bundestagsbefragung im Rahmen der Leopoldina-Evidenzinitiative unter Leitung von Claudia Buch, Regina T. Riphahn und Monika Schnitzer*. Tech. rep. MyCoRe Community, vol. 27.

Chu, Jonathan A. and Stefano Recchia (2022). “Does Public Opinion Affect the Preferences of Foreign Policy Leaders? Experimental Evidence from the UK Parliament”. *The Journal of Politics* 84.3, pp. 1874–1877.

Deutscher Bundestag (2022). *Rules of Procedure of the German Bundestag and Rules of Procedure of the Mediation Committee*.

Fagan, E. J. and Zachary A. McGee (2022). “Problem Solving and the Demand for Expert Information in Congress”. *Legislative Studies Quarterly* 47.1, pp. 53–77.

Göbel, Sascha and Simon Munzert (2022). “The Comparative Legislators Database”. *British Journal of Political Science* 52.3, pp. 1398–1408.

Lee, Nathan (2022). “Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers”. *American Political Science Review* 116.2, pp. 677–688.

Loewen, Peter John, Daniel Rubenson, and John R. McAndrews (2022). “When Do Politicians Pursue More Policy Information?” *Journal of Experimental Political Science* 9.2, pp. 216–224.

Pereira, Miguel M. (2022). “How do Public Officials Learn About Policy? A Field Experiment on Policy Diffusion”. *British Journal of Political Science* 52.3, pp. 1428–1435.

Priem, Jason, Heather Piwowar, and Richard Orr (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*.

Riphahn, Regina T. and Monika Schnitzer (2022). "Die Bedeutung wissenschaftlicher Evidenz für die Arbeit des Bundestags". *Wirtschaftsdienst* 102.4, pp. 270–272.

Ban, Pamela, Ju Yeon Park, and Hye Young You (2023). "How Are Politicians Informed? Witnesses and Information Provision in Congress". *American Political Science Review* 117.1, pp. 122–139.

Bellodi, Luca (2023). "Bureaucratic Information in Congress".

Ouimet, Mathieu, Morgane Beaumier, Adrien Cloutier, Alexandre Côté, Éric Montigny, François Gélineau, Steve Jacob, and Stéphane Ratté (2023). "Use of research evidence in legislatures: a systematic review". *Evidence & Policy* -1.aop, pp. 1–18.

Park, Ju Yeon (2023). "Electoral rewards for political grandstanding". *Proceedings of the National Academy of Sciences* 120.17, e2214697120.

Rasmussen, Anne and Stefanie Reher (2023). "(Inequality in) Interest Group Involvement and the Legitimacy of Policy Making". *British Journal of Political Science* 53.1, pp. 45–64.

Vallejo Vera, Sebastián (2023). "By invitation only: on why do politicians bring interest groups into committees". *The Journal of Legislative Studies* 29.1, pp. 1–38.

Ban, Pamela, Ju Yeon Park, and Hye Young You (2024). *Hearings on the Hill: The Politics of Informing Congress*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press.

Coil, Collin, Caroline Bruckner, Natalie Williamson, Karen O'Connor, and Jeff Gill (2024). "Still Underrepresented? Gender Representation of Witnesses at House and Senate Committee Hearings". *Journal of Women, Politics & Policy*.

Geddes, Marc (2024). "Comparing evidence use in parliaments: the interplay of beliefs, traditions, and practices in the UK and Germany". *Policy and Society* 43.4, pp. 447–462.

Hünermund, Sebastian (2024). *Öffentliche Anhörungen im Deutschen Bundestag: Eine empirische Untersuchung ihrer Funktionen im politischen Prozess*. 1. Auflage. Studien zum Parlamentarismus 37. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG.

Senninger, Roman and Henrik B. Seeberg (2024). "Which Information Do Politicians Pay Attention To? Evidence from a Field Experiment and Interviews". *British Journal of Political Science*, pp. 1–18.

Willems, Evelien, Bart Maes, and Stefaan Walgrave (2024). "Mechanisms of Political Responsiveness: The Information Sources Shaping Elected Representatives' Policy Actions". *Political Research Quarterly* 77.3, pp. 851–865.

Ban, Pamela, Ju Yeon Park, and Hye Young You (2025). "Bureaucrats in Congress: The Politics of Inter-branch Information Sharing". *The Journal of Politics*, p. 735440.

Ramirez-Ruiz, Sebastian (2025). *Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience*.

Rasmussen, Anne and Stefanie Reher (2025). "Public opinion towards interest groups: The differential impact of ties to cause and business groups". *European Journal of Political Research* 64.2, pp. 900–914.

Vivalt, Eva, Aidan Coville, and Sampada Kc (2025). "Local knowledge, formal evidence, and policy decisions". *Journal of Development Economics* 174, p. 103425.

The Bundestag Expert Witness Tracker (BEWIT): A database of German Bundestag public expert hearings

Online Appendix

Contents

Appendix A Variable Codebook

1. Hearings Stores metadata for each Bundestag expert hearing.

- `mt_id`: Unique identifier for each hearing (primary key).
- `mt_date`: Date on which the hearing took place.
- `mt_parlperiod`: Parliamentary period (legislative term) in which the hearing occurred.
- `mt_title`: Title of the hearing, typically reflecting the main topic.
- `mt_committee`: Name of the parliamentary committee responsible for the hearing.
- `mt_chair`: Chairperson of the committee during the hearing.
- `mt_motiongroup`: Political motion group (e.g., party or parties) that initiated the hearing.
- `mt_protocol`: Permanent hyperlink to the official verbatim transcript of the hearing.
- `mt_files_disc`: Identifiers of associated documents or legislative files discussed.

2. List of experts Central table linking experts to hearings and affiliations.

- `mt_id`: Foreign key linking to the associated hearing (`Hearings.mt_id`).
- `exp_aff_id`: Unique expert-affiliation identifier (linked to Academics table).
- `exp_name`: Full name of the expert witness.
- `exp_title`: Professional title or academic degree (e.g., Prof., Dr.).
- `affil_id`: Foreign key linking to the affiliation (`Affiliations.affil_id`).
- `affil_name`: Name of the affiliated institution or organization.
- `affil_sector`: Sector classification of the affiliation (e.g., research, government).

3. Academics Contains academic profiles for experts with identifiable research careers.

- `openalex_id`: Persistent OpenAlex identifier for the researcher.
- `sci_field`: Broad scientific field (e.g., Social science).
- `sci_discipline`: Specific scientific discipline (e.g., Political science, Law).
- `sci_last_inst`: Most recent institutional affiliation in OpenAlex.
- `exp_aff_id`: Foreign key linking to expert-affiliation pair (`List of Experts.exp_aff_id`).
- `exp_name`: Name of the expert (matching `List of Experts`).

4. Affiliations Provides metadata about institutions and organizations represented in hearings.

- `affil_id`: Unique identifier for each affiliation (primary key).
- `affil_name`: Full name of the organization or institution.
- `affil_sector`: Sector classification (e.g., academia, civil society, public sector).
- `lobbyreg_id`: Bundestag transparency register ID, if applicable.

Appendix B Supporting Figures and Tables

Characteristic	OR	95% CI	p-value
Change in reporting protocol	0.76	0.65, 0.89	<0.001
Committee			
Budget	—	—	
Climate Action and Energy	11.2	5.52, 26.0	<0.001
Cultural and Media Affairs	7.79	3.35, 20.0	<0.001
Digital Affairs	4.35	2.03, 10.5	<0.001
Economic Affairs	5.03	2.29, 12.3	<0.001
Economic Cooperation and Development	7.45	2.78, 21.5	<0.001
Educ, Research and Tech Assessment	6.07	2.78, 14.8	<0.001
Env, Nature Cons, Nuclear Safety and Consumer Prot	11.5	5.39, 27.7	<0.001
European Union Affairs	0.42	0.02, 2.54	0.4
Family Affairs, Senior Citizens, Women and Youth	22.2	10.3, 53.8	<0.001
Finance	5.59	2.67, 13.2	<0.001
Food and Agriculture	4.25	1.80, 11.0	0.002
Foreign Affairs	2.23	0.80, 6.36	0.13
Health	3.13	1.50, 7.41	0.005
Housing, Urb Dev, Building and Local Gov	16.8	7.84, 40.4	<0.001
Human Rights and Humanitarian Aid	9.73	4.21, 24.9	<0.001
Internal Affairs and Community	4.78	2.33, 11.2	<0.001
Labor and Social Affairs	12.9	6.26, 30.4	<0.001
Legal Affairs	7.79	3.85, 18.0	<0.001
Main Committee	3.43	1.22, 10.1	0.021
Sports	9.36	3.82, 25.2	<0.001
Tourism	8.10	3.85, 19.2	<0.001
Transport	16.9	7.96, 40.5	<0.001
N. observations	3,594		

Table B1: Logistic regression estimates of interest group representation in expert hearings. Odds ratios (OR), 95% confidence intervals, and p-values from a logistic regression predicting whether an expert witness represents special interests. The model includes a binary indicator for a change in the reporting protocol (transparency reform) and fixed effects for parliamentary committees. The reference committee category is *Budget*. Results suggest a statistically significant association between the reform and a lower likelihood of interest group representation, and large between-committee differences in baseline probabilities.

	OpenAlex ID			Lobbyregister ID		
Characteristic	OR	95% CI	p-value	OR	95% CI	p-value
Year						
2010	—	—		—	—	
2011	1.05	0.48, 2.20	0.9	0.90	0.69, 1.16	0.4
2012	2.18	0.96, 4.90	0.058	0.81	0.62, 1.07	0.13
2013	1.02	0.45, 2.25	>0.9	0.87	0.67, 1.14	0.3
2014	1.64	0.65, 4.34	0.3	1.07	0.80, 1.44	0.6
2015	1.52	0.65, 3.54	0.3	0.95	0.74, 1.23	0.7
2016	0.98	0.45, 2.09	>0.9	1.00	0.78, 1.28	>0.9
2017	1.44	0.64, 3.20	0.4	0.99	0.76, 1.31	>0.9
2018	1.22	0.55, 2.63	0.6	0.81	0.61, 1.09	0.2
2019	0.76	0.37, 1.47	0.4	0.96	0.75, 1.23	0.7
2020	2.30	1.05, 4.86	0.031	1.07	0.84, 1.37	0.6
2021	1.11	0.54, 2.14	0.8	1.00	0.79, 1.28	>0.9
2022	2.24	0.99, 4.94	0.047	1.13	0.88, 1.44	0.3
2023	1.73	0.81, 3.50	0.14	1.11	0.87, 1.41	0.4
2024	2.08	0.92, 4.59	0.071	1.11	0.85, 1.44	0.4
N. observations	3,512			10,474		

Table B2: Logistic regression estimates of linkage success by year. The table presents odds ratios (OR), 95% confidence intervals, and p-values from logistic regression models estimating the likelihood of a successful match to an OpenAlex author ID or a Lobbyregister organization ID as a function of year. The reference year is 2010.

Invitation source	N.	Percent
CDU/CSU	621	25.6%
SPD	587	24.2%
BÜNDNIS 90/DIE GRÜNEN	356	14.7%
FDP	259	10.7%
All groups	187	7.7%
DIE LINKE	147	6.1%
Special participation rights	124	5.1%
AfD	78	3.2%
Other*	60	2.5%
Gruppe BSW	5	0.2%

Table B3: Distribution of experts invited by each group. **Other* refers to combinations of Chair and a mix of parties.

Appendix C Software statement

I used R version 4.4.2 (R Core Team, 2024) and the following R packages:

`broom` v. 1.0.7 (Robinson, Hayes, and Couch, 2024)
`DBI` v. 1.2.3 (R Special Interest Group on Databases (R-SIG-DB), Wickham, and Müller, 2024)
`dbplyr` v. 2.5.0 (Wickham, Girlich, and Ruiz, 2024)
`dplyr` v. 1.1.4 (Wickham, François, et al., 2023)
`forcats` v. 1.0.0 (Wickham, 2023a)
`ggplot2` (Wickham, 2016)
`ggtext` v. 0.1.2 (Wilke and Wiernik, 2022)
`glue` v. 1.8.0 (Hester and Bryan, 2024)
`gt` v. 0.11.1 (Iannone et al., 2024)
`gtsummary` (Sjoberg et al., 2021)
`httr` v. 1.4.7 (Wickham, 2023b)
`janitor` v. 2.2.0 (Firke, 2023)
`jsonlite` (Ooms, 2014)
`lubridate` (Grolemund and Wickham, 2011)
`marginaleffects` (Arel-Bundock, Greifer, and Heiss, 2024)
`openalexR` (Massimo et al., 2024)
`purrr` v. 1.0.4 (Wickham and Henry, 2025)
`readxl` v. 1.4.3 (Wickham and Bryan, 2023)
`rvest` v. 1.0.4 (Wickham, 2024)
`shadowtext` v. 0.1.4 (Yu, 2024)
`stringr` v. 1.5.1 (Wickham, 2023c)
`tidyr` v. 1.3.1 (Wickham, Vaughan, and Girlich, 2024)
`webshot2` v. 0.1.1 (Chang, 2023)
`writexl` v. 1.5.1 (Ooms, 2024)
`xtable` v. 1.8-4 (Dahl et al., 2019)

References

Grolemund, Garrett and Hadley Wickham (2011). “Dates and Times Made Easy with lubridate”. *Journal of Statistical Software* 40.3, pp. 1–25.

Ooms, Jeroen (2014). “The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects”. *arXiv:1403.2805 [stat.CO]*.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton (2019). *xtable: Export Tables to LaTeX or HTML*.

Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange (2021). “Reproducible Summary Tables with the gtsummary Package”. *The R Journal* 13 (1), pp. 570–580.

Wilke, Claus O. and Brenton M. Wiernik (2022). *ggtext: Improved Text Rendering Support for 'ggplot2'*.

Chang, Winston (2023). *webshot2: Take Screenshots of Web Pages*.

Firke, Sam (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*.

Wickham, Hadley (2023a). *forcats: Tools for Working with Categorical Variables (Factors)*.

— (2023b). *httr: Tools for Working with URLs and HTTP*.

— (2023c). *stringr: Simple, Consistent Wrappers for Common String Operations*.

Wickham, Hadley and Jennifer Bryan (2023). *readxl: Read Excel Files*.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan (2023). *dplyr: A Grammar of Data Manipulation*.

Arel-Bundock, Vincent, Noah Greifer, and Andrew Heiss (2024). “How to Interpret Statistical Models Using marginaleffects for R and Python”. *Journal of Statistical Software* 111.9, pp. 1–32.

Hester, Jim and Jennifer Bryan (2024). *glue: Interpreted String Literals*.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy (2024). *gt: Easily Create Presentation-Ready Display Tables*.

Massimo, Aria, Le Trang, Cuccurullo Corrado, Belfiore Alessandra, and Choe June (2024). “openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex”. *The R Journal* 15 (4), pp. 167–180.

Ooms, Jeroen (2024). *writexl: Export Data Frames to Excel 'xlsx' Format*.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham, and Kirill Müller (2024). *DBI: R Database Interface*.

Robinson, David, Alex Hayes, and Simon Couch (2024). *broom: Convert Statistical Objects into Tidy Tibbles*.

Wickham, Hadley (2024). *rvest: Easily Harvest (Scrape) Web Pages*.

Wickham, Hadley, Maximilian Girlich, and Edgar Ruiz (2024). *dbplyr: A 'dplyr' Back End for Databases*.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich (2024). *tidyr: Tidy Messy Data*.

Yu, Guangchuang (2024). *shadowtext: Shadow Text Grob and Layer*.

Wickham, Hadley and Lionel Henry (2025). *purrr: Functional Programming Tools*.

Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory

Sebastian Ramirez-Ruiz (Hertie School)*
Simon Munzert (Hertie School)

Abstract. Being able to make sense of scientific evidence is a critical component in shaping effective public policy, enabling both policymakers and the public to make informed decisions on complex issues. However, existing measures of this skillset may not fully address the specific methodological and applied demands of public policy contexts. We develop the *Inventory for Numeracy, Statistics, and Policy-oriented Inference and REasoning* (INSPIRE), a comprehensive measurement instrument to assess competence in scientific evidence consumption in relation to public policy. The instrument integrates knowledge of statistical reasoning, data and visualization literacy, causal reasoning and inference, and the scientific method, alongside the ability to critically evaluate scientific information pertinent to policy debates. Using Item Response Theory (IRT), we systematically assess an initial item pool, resulting in an inventory of 30 items. We assessed the inventory's psychometric and substantive validity across three samples: a general population, policy students and professionals in a data science training program, and participants in a pre-election forecasting study. This allowed us to examine its performance across different groups and contexts. Results indicate good internal consistency and evidence of construct, criterion, and predictive validity. We close with considerations for potential use of the item bank in research and applied settings.

Keywords. scientific evidence consumption | scientific reasoning | public policy | evidence-informed policymaking | item response theory (IRT) | Test instrument

*Corresponding author. Address: Friedrichstrasse 180, 10117 Berlin, Germany. Email: ramirez-ruiz@hertie-school.org, Web: seramirezruiz.github.io.

Introduction

Current policy challenges, such as climate change, AI regulation, and food security, underscore the vital role that scientific insights play in shaping the policy debate. These issues require complex responses to balance competing interests, mitigate risks, and achieve sustainable results. Recognizing this need, many national and supranational entities increasingly establish governance frameworks that promote evidence-informed decision-making (Juncker, 2014; U.S. Congress, 2019; Juncker, 2014; U.S. Congress, 2019). Using scientific evidence to inform policymaking presupposes that policymakers and public officials have a foundational understanding of scientific concepts, enabling them to translate scientific insights into practical policy solutions. In other words, an informed policy workforce is essential for ensuring that evidence-informed frameworks are not just aspirational but are effectively implemented.

In the general public arena, however, the picture is more complex. While there is broad support for integrating science into policymaking—reflected in global survey data showing that a majority of respondents agree scientists should actively participate in political debates to inform policy decisions (European Commission, 2021; Cologna et al., 2025; European Commission, 2021; Cologna et al., 2025)—this sentiment is not universally shared. Growing tensions persist, with segments of the population expressing declining trust in both scientific institutions and democratic processes (Tyson and Kennedy, 2024). These divisions present a crucial challenge: ensuring that policies grounded in evidence maintain public legitimacy. A scientifically competent policy workforce is essential not only for making informed decisions but also for bridging legitimacy gaps that arise when parts of the public perceive a disconnect between scientific evidence and policy outcomes (Rimkutė and Mazepus, 2025).

Despite the importance of the ability to engage with scientific evidence in policy contexts, fields concerned with policy practice have paid little attention to systematically measuring it. Existing tools, often adapted from more general educational or cognitive frameworks, might fall short when applied to public policy. While these measures evaluate general scientific knowledge or abstract reasoning skills, they may not address the specific demands of policy decision-making. Effective policymaking often requires competencies relating to numeric and statistical reasoning, causal inference, and the ability to critically evaluate evolving evidence from real-world scenarios. This gap underscores the need for a comprehensive approach to measuring scientific evidence consumption literacy in a way that directly relates to public policy.

In this paper, we introduce the *Inventory for Numeracy, Statistics, and Policy-oriented Inference and REasoning* (INSPIRE), a toolset designed to assess essential competencies in scientific evidence consumption within the policy domain. INSPIRE is grounded in Item Response Theory (IRT), which we use to validate a fixed-form inventory that balances measurement precision with respondent efficiency. We validate the instrument rigorously across three distinct samples: a diverse online respondent pool including government workers, policy students and professionals engaged in data science training, and participants from a pre-election forecasting study. The instrument demonstrates strong internal consistency and the capacity to distinguish between theoretical knowledge and applied reasoning skills. Our results support INSPIRE as a psychometrically robust and practically scalable tool for evaluating scientific evidence consumption skills in policy-relevant settings.

To ensure broad applicability, we provide a validated item bank of 30 items along with recommendations for its flexible implementation in different research and policy contexts. This allows for the construction of inventories tailored to specific needs, such as quick assessments, experimental designs, or training interventions aimed at improving scientific reasoning skills.

A tailored measurement tool for scientific reasoning in public policy holds significant value for both researchers and policy practitioners. As an explanatory variable, such a measure could deepen our understanding of critical phenomena, such as policymakers' receptivity to impact evaluations (Toma and Bell, 2024) or their ability to update beliefs and decisions in response to new evidence (Vivalt and Coville, 2023). At the same time, it can serve as a meaningful outcome measure, reflecting a growing interest in enhancing these skills among policymakers (Crowley et al., 2021; Mehmood, Naseer, and Chen, 2024; Crowley et al., 2021; Mehmood, Naseer, and Chen, 2024). As such, it provides a device for assessing the effectiveness of interventions designed to strengthen policymakers' capacities to consume, evaluate, and apply evidence. By linking measurement with actionable insights, this tool can help bridge the gap between aspiration and implementation, ultimately contributing to a more informed and effective policy workforce.

Why evidence consumption skills matter for public policy

The importance of scientific competence extends beyond the realm of policy and is deeply embedded in educational systems worldwide. Many early-stage curricula are designed to equip students with the scientific concepts and methods necessary to navigate complex issues in daily life. At the heart of these efforts lies, scientific reasoning—conceived as "*the ability to evaluate the quality of scientific evidence*" (Drummond and Fischhoff, 2017a). This capacity is not only foundational to science education but is also regarded as essential for the functioning of modern societies (Economic Co-operation and Development, 2006; DeBoer, 2011; Economic Co-operation and Development, 2006; DeBoer, 2011).

Research on scientific literacy and reasoning has deep roots in developmental psychology and has been implicitly present in influential models of rationality (e.g., Inhelder and Piaget, 1958). This historical context has shaped much of the related concepts' methodological landscape, leading to a primary emphasis on the education sciences—both in terms of measurement development and the design of interventions to enhance reasoning skills (Fischer et al., 2014, for an overview). Consequently, while competences related to scientific evidence consumption are broadly relevant, much of the research in this domain has been centered on how children and adolescents acquire and refine these skills, often within structured educational environments.

Beyond the classroom, however, there is growing recognition of the broader societal implications of a scientifically competent public. A body of research has explored how laypeople's ability to evaluate scientific evidence influences civic engagement, public discourse, and trust in science. Some evidence suggests that scientific reasoning and knowledge might be related to the entrenchment of polarized beliefs on contentious science topics (Drummond and Fischhoff, 2017b) and increased susceptibility to misinformation (Fazio et al., 2015; Linden, 2022; Fazio et al., 2015; Linden, 2022). This highlights a critical issue: scientific evidence consumption literacy is not just an academic or educational concern but a skill set with real-world consequences for democratic decision-making, public policy, and societal resilience against misinformation. Given these stakes, efforts to measure and improve how individuals engage with scientific findings must extend beyond formal education settings and consider populations that wield influence over public affairs—such as policymakers.

Despite the recognized importance of competences in scientific evidence consumption, existing measurement tools have key limitations in the policy context. In their comprehensive review, Opitz, Heene, and Fischer (2017) identified 38 instruments designed to assess skills under the umbrella of scientific reasoning. However, they highlight critical gaps. First, most were designed for educational settings, particularly K-12 students, and rarely extend beyond these contexts. Additionally, many of these tools focus on natural and applied sciences, whereas policymakers more often encounter and rely on evidence from the social sciences (Ramirez-Ruiz and Senninger, 2025). This contextual and disciplinary imbalance highlights a crucial limitation of existing measures, which may not adequately capture the type of evidence that policymakers most frequently encounter when interpreting scientific evidence in their work.

Scientific evidence consumption in public policy involves distinct challenges and demands that must be addressed. Unlike students in structured learning environments, policymakers must consume scientific insights within high-stakes decision-making processes, where the consequences extend beyond individual learning outcomes to public welfare, economic stability, and societal well-being. In this context, we are concerned not with individuals acting as basic researchers but with those who take on the role of scientifically knowledgeable practitioners—professionals who must interpret and apply existing scientific knowledge rather than generate new discoveries. These decisions require not only technical understanding but also the ability to evaluate diverse evidence, assess causal relationships, and make informed judgments under uncertainty. Whereas traditional assessments of scientific competencies often emphasize knowledge generation, effective policymaking relies more heavily on evaluating evidence and drawing conclusions in applied contexts, often under time constraints and in the face of competing interests.

Additionally, the constraints faced by policymakers—such as limited time, attention, and competing demands—pose logistical challenges for both the assessment of these skills and the engagement of respondents. Given these realities, any measure of scientific evidence consumption for public policy must be designed to balance methodological rigor with practical feasibility. To be useful in real-world policy settings, assessment tools must be efficient, adaptive, and capable of capturing relevant competencies without imposing undue burden on respondents. This paper seeks to address these gaps by developing a scientifically grounded yet practically viable inventory of items tailored to the unique demands of the policy world.

Conceptual framework

We start with a loose definition of the concept of interest: *Scientific evidence consumption literacy* is the ability to make sense of scientific research and data in everyday life. It means understanding the basics of how science works, discerning reliable from less reliable evidence, spotting trustworthy sources, and being able to reason about numbers, statistics, and cause-and-effect relationships. It also includes knowing how to read charts and graphs without getting fooled. As such, evidence consumption literacy has both a knowledge and a skill component.

The term "evidence" requires further specification within the scope of this study. Here, we primarily focus on types of evidence that policymakers are likely to encounter in a quantitative format. Specifically, our emphasis is on empirical, quantitative evidence rather than qualitative evidence. This focus is motivated, first, by the alignment with existing measures of statistical and data literacy; second, by the relative practicality of assessment compared to, for example, the analytical skills required to engage with qualitative evidence; and third, by the assumption that these quantitative skills may present a higher barrier for policymakers without formal training in empirical sciences (including the social, medical, and natural sciences), where such methods and forms of evidence are more commonly taught and applied. That said, fundamental competencies in scientific reasoning and causal inference—both of which are components of our measure—are equally relevant for evaluating qualitative evidence.

Scientific reasoning is widely regarded as an umbrella concept encompassing multiple cognitive and analytical skills. Previous research has sought to define and measure these competencies by focusing on distinct but interrelated components, including hypothesis generation, evidence generation, evidence evaluation, and conclusion drawing (Opitz, Heene, and Fischer, 2017). These processes are fundamental to scientific inquiry and have been the basis of many existing assessment tools.

Building on this foundation, Fischer et al. (2014) propose a broader conceptualization of scientific reasoning that integrates epistemic modes of engagement with distinct reasoning activities. They identify three primary modes of engagement that shape how individuals interact with scientific information. The first is *theory-building about phenomena*, which involves engaging in conceptual reasoning to contribute to knowledge creation. The second is *science-based reasoning and argumentation in practice*, where individuals apply scientific reasoning to assess and justify decisions in real-world contexts. The third mode is *artifact-centered reasoning*, which focuses on evaluating whether and how a given problem can be addressed using scientific tools or technological solutions. These modes, in turn, imply a set of core 'epistemic' activities, which include: Identifying problems, questioning, generating hypotheses, constructing and redesigning artifacts, generating evidence, evaluating evidence, drawing conclusions, and communicating and scrutinizing scientific claims.

Our approach to developing this inventory was guided by these conceptual distinctions while adapting them to the specific demands of public policy decision-making. While all three modes of engagement are relevant, we argue that science-based reasoning in practice (Mode 2) and artifact-centered reasoning (Mode 3) are the most applicable. Rather than engaging in original theory-building and knowledge creation, policy professionals primarily use scientific evidence consumption to assess whether scientific constructs, empirical findings, or technological solutions are relevant to specific policy challenges.

Furthermore, we integrate these ideas with core competencies from inter-related domains, including scientific literacy, statistical literacy, causal reasoning, and data literacy. Each of these fields has developed structured frameworks for assessing reasoning skills in applied contexts, often highlighting their relevance for sound decision-making in both personal and professional domains. We build on their insights to ensure that our inventory reflects both the conceptual dimensions of scientific evidence consumption literacy in policy work, as well as their practical applications in the kinds of decisions faced by policy professionals.

Scientific literacy

Scientific literacy refers to the "knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity" (National Academies of Sciences, Engineering, and Medicine, 1996). In what might be one of the most influential pieces in this space, Miller (1983) highlights three key dimensions to scientific literacy: a) understanding the scientific approach (i.e., the norms and methods of science), b) understanding basic scientific constructs (i.e., basic language and concepts), and c) understanding of science policy issues (i.e., awareness of the impact of science and technology on society).

Measurement efforts in this domain reflect different emphases on these dimensions. For instance, the Test of Scientific Literacy Skills (TOSLS) (Gormally, Brickman, and Lutz, 2012) assesses college-level students' ability to reason with scientific evidence. It focuses on two primary skill areas: (1) recognizing and analyzing how scientific knowledge is generated through methods of inquiry, and (2) organizing, interpreting, and evaluating scientific data and information. Other approaches can be seen, for example in the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR) (Romine, Sadler, and Kinslow, 2017), which frames scientific literacy through the lens of engagement with real-world, socially relevant science issues. It evaluates four domains: complexity (recognizing multiple factors in science-related problems), perspective-taking, inquiry (understanding how science informs answers), and skepticism (critical stance toward claims).

Scientific literacy serves as a crucial foundational concept for our purposes, as it encompasses the knowledge of the scientific approach, including its core language and concepts, which are essential for evaluating scientific evidence. For example, a policymaker who understands the notion of scientific consensus—not as unanimous agreement, but as a convergence of evidence supporting a particular hypothesis—is better equipped to make sense of scientific findings on politically contested issues such as climate change or public health. Similarly, grasping the nature of science as a method of establishing truth based on empirical evidence, rather than authority or ideology, helps policymakers avoid conflating scientific input with mere opinion.

In designing our inventory, we adopt and adapt several of these conceptual components. Specifically, we center on items that assess foundational reasoning about scientific processes and the credibility of evidence. However, while scientific literacy is important, existing measures tend to fail to capture the practical skills required to engage with data and evidence in real-world formats. This is why we expand our focus to include additional competencies.

Statistical literacy

Statistical literacy, like scientific literacy, is increasingly framed not only as a technical competency but also as a foundational civic skill. According to Wallman (1993), it involves "the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions."

Efforts to define and measure statistical literacy reflect its cognitive complexity. For example, Watson and Callingham (2003) conceptualize statistical literacy as a hierarchical construct involving progressively sophisticated reasoning processes. Their framework captures a range of competencies—from interpreting basic statistical information and proportional reasoning to evaluating uncertainty and critically engaging with the context in which statistics are used. These layers emphasize that statistical thinking involves not just computation but judgment. Other instruments, such as the Take CARE (Context, Assembly, Randomness and Error) (Schield, 2010) and the REALI (Reasoning about Everyday and Academic Literacy Items) (Sabbag, Garfield, and Zieffler, 2018), offer a more functional view of statistical literacy, focusing on how well individuals can recall definitions, interpret data representations, and understand foundational concepts such as variability, probability, and hypothesis testing. While less focused on complex inferential reasoning, such frameworks underscore the importance of core interpretive skills in applied contexts.

In our inventory, we draw on these conceptual takes to develop items targeting statistical reasoning. These items assess the ability to interpret statistical information, reason about uncertainty and spread, and evaluate claims derived from statistical models in policy-relevant settings.

Machine learning and related statistical tools—often broadly labeled as "AI technologies"—are playing an increasingly prominent role in policy contexts. From drafting legislation to regulate the use of AI in sectors like healthcare or education, to deploying machine learning systems in public administration for tasks such as resource allocation or fraud detection, policymakers are routinely confronted with decisions that hinge on a basic understanding of these technologies. For instance, evaluating the risks and benefits of predictive policing tools requires more than general statistical literacy; it demands a grasp of how machine learning models are trained, validated, and where their limitations lie. While a few recent scales attempt to measure AI and machine learning literacy (Hornberger, Bewersdorff, and Nerdel, 2023; Wang, Rau, and Yuan, 2023; Weber, Pinski, and Baum, 2023; Ng et al., 2024; Hornberger, Bewersdorff, and Nerdel, 2023; Wang, Rau, and Yuan, 2023; Weber, Pinski, and Baum, 2023; Ng et al., 2024), they are typically designed for technical or educational settings and do not align with the needs of policy professionals, often being too specialized and assuming too much background knowledge. To address this gap, we extend our assessment by including a set of items that capture fundamental knowledge of machine learning: what it does, how it is evaluated, and where its risks and limitations are most salient in a policy context.

Data and visualization literacy

Policymakers are routinely presented with data in the form of graphs, charts, and dashboards, but also through summaries such as probabilities, percentages, and risk estimates. Without a solid grounding in data and visualization literacy, they may misjudge trends, misunderstand uncertainty, or draw incorrect conclusions from seemingly straightforward figures. For example, interpreting a 10% increase in unemployment or a 90% vaccine efficacy rate requires not only statistical understanding but also the ability to contextualize these numbers in terms of real-world impact. Being able to critically engage with how data is presented—both visually and numerically—is therefore essential for translating evidence into sound policy.

Data literacy has some overlaps with statistical literacy, yet it introduces distinct emphases—particularly around working with, communicating, and reasoning through data in applied and visual formats. D'Ignazio and Bhargava (2016) outline four key dimensions of data literacy: reading data (understanding what data represents), working with data (collecting, cleaning, managing), analyzing data (sorting, aggregating, and comparing), and arguing with data (building a compelling narrative using data as evidence). This framework highlights not just technical fluency, but also the communicative and contextual nature of data work. There are further conceptualizations expanding on this foundation. For instance, Wolff et al. (2016) define data literacy as the ability to ask and answer real-world questions using data through an inquiry-driven process, with attention to ethical use. Their view centers on both technical competencies—such as selecting, cleaning, visualizing, and interpreting data—and creative and design-oriented skills for communicating stories and insights. Importantly, this framing also emphasizes the role of media and visual literacy in understanding and conveying data.

A related and emerging area is data visualization literacy, formalized as its own construct. Börner, Bueckle, and Ginda (2019) propose a framework that encompasses understanding the communication goals of a visualization, interpreting different types of charts, recognizing variables and symbols, and navigating levels of scale, interaction, and visual encoding strategies. Some efforts to measure these skills build on cognitive models of how users engage with graphs and visual data. For example, Boy et al. (2014) describe a sequential information extraction process—from setting a goal and parsing patterns to comparing visual elements and drawing inferences. Another approach is the Visualization Literacy Assessment Test (VLAT) (Lee, Kim, and Kwon, 2017), which assesses users' ability to perform specific tasks (e.g., retrieve values, detect trends, compare distributions) across a range of graphical formats.

More recently, attention has shifted toward the critical evaluation of visualizations, especially in the context of misinformation. The Critical Thinking Assessment for Literacy in Visualizations (CALVI) (Ge, Cui, and Kay, 2023) develops a measure specifically focused on users' ability to detect and reason about 'misleaders'—what they define as visual design choices that can bias interpretation. The authors argue that traditional visualization literacy assessments often assume well-designed, truthful graphics, overlooking the fact that not all visualizations are constructed with clarity or integrity in mind. The CALVI framework defines a space of misleading strategies (e.g., inappropriate axes, selective data inclusion) and provides validated assessment items to evaluate participants' ability to evaluate visual evidence critically and avoid being misled.

Drawing on these ideas, our inventory includes a set of items designed to assess proficiency in handling numerical data and interpreting visual information. These tasks target practical reasoning skills necessary for using data effectively in policy communication and decision-making.

Causal reasoning

Causal reasoning refers to the cognitive and analytical processes used to identify, evaluate, and explain cause-and-effect relationships. It spans both descriptive approaches—how people intuitively make causal judgments—and more formal approaches—how such reasoning should be structured to support valid inference (Waldmann, 2017).

From an inference standpoint, Judea Pearl's work has been foundational in formalizing causal reasoning through structural causal models (SCMs) and graphical representations. He offers a helpful framework through his "ladder of causation" (Pearl and Mackenzie, 2018), outlining three levels of causal understanding: (1) association (seeing correlations), (2) intervention (understanding what happens if we do something), and (3) counterfactuals (imagining what would have happened under different circumstances). Pearl's framework emphasizes that causal inference requires more than statistical association—it requires assumptions, models, and a capacity for counterfactual reasoning.

In applied fields such as policy analysis, causal reasoning is essential for evaluating the effectiveness of interventions and drawing appropriate conclusions from empirical evidence. The Neyman-Rubin potential outcomes framework has become a cornerstone for estimating causal effects in practice (Angrist and Pischke, 2009; Imbens and Rubin, 2015; Angrist and Pischke, 2009; Imbens and Rubin, 2015). Yet the conceptual foundations of these models—such as understanding counterfactuals, identifying sources of bias, and recognizing the limits of causal claims—are often poorly understood outside specialized training.

While causal inference is a central concern in policy research, there are no assessment tools aimed at measuring individuals' ability to reason causally in applied or policy-relevant contexts. Existing approaches tend to be embedded in methodological training and inferred from behavioral cues (i.e., did policymakers act in accordance with what we would expect) (Mehmood, Naseer, and Chen, 2024), rather than in assessments of features of causal reasoning.

Guided by these conceptual frameworks and the absence of established measurement tools, we developed a set of inventory items designed to capture key aspects of causal reasoning in policy settings. These items tackle individuals' ability across the rungs of the "ladder of causation"—distinguishing correlation from causation, evaluating research designs and evidence, applying counterfactual reasoning, and more generally, grappling with the complexities inherent in causal inference.

Inventory development

Based on this conceptual foundation, we developed a 52-item test inventory covering four key areas related to scientific evidence consumption literacy in the policy domain. While we group items under distinct thematic baskets, we recognize that these categories are not rigidly bounded. In practice, competencies such as data interpretation, scientific reasoning, and causal inference often intersect. Our groupings reflect our primary logic rather than a very strict conceptual separation:

- **Scientific literacy (SEL)** (15 items): Assesses foundational knowledge of scientific processes, including familiarity with peer review, scientific consensus, research design, and best practices for evaluating evidence reliability in policy contexts.
- **Statistical literacy (SML)** (14 items): Evaluates high-order statistical reasoning, including the ability to interpret statistical information, uncertainty, and critically evaluate machine learning models and their implications for policy decisions.
- **Data and visualization literacy (DVL)** (10 items): Measures proficiency in working with numerical data, performing basic calculations, and interpreting visualizations to support policy communication and decision-making.
- **Causal reasoning (CR)** (13 items): Examines the ability to distinguish correlation from causation, assess research designs for causal inference, evaluate causal claims, understand the challenges of establishing cause-effect relationships in policy evaluation, apply counterfactual reasoning, and determine whether a body of evidence supports a causal conclusion.

We designed the initial test inventory through a multi-step process combining literature review, targeted adaptation of existing measures, and small-scale piloting. Our goal was to create items that reflect the real-world demands of scientific evidence consumption and reasoning in public policy contexts, while maintaining conceptual alignment with our core constructs. The resulting tool was designed to be a) psychometrically sound, b) contextually relevant, and c) practical to administer in time-constrained policy settings.

We began by reviewing existing assessments of scientific reasoning and quantitative literacy. Our measurements build on and adapt items from validated instruments such as the Test of Scientific Literacy Skills (TOSLS) (Gormally, Brickman, and Lutz, 2012), the Critical Thinking Assessment for Literacy in Visualizations (CALVI) (Ge, Cui, and Kay, 2023), and the Scientific Reasoning Scale (Drummond and Fischhoff, 2017a). Additionally, we generated new items aligned with four conceptual subdomains: scientific literacy, statistical literacy, causal reasoning, and data and visualization literacy.

Most items were newly authored to fit this framework, though we also adapted select items from existing sources to better reflect policy-relevant scenarios. For instance, one original TOSLS item asked respondents to compare the statistical uncertainty of two studies estimating the caffeine content of an energy drink. We retained the core structure of the question but modified its content to align with public policy domains. In our version, the item presented two studies estimating the average educational level in two rural regions, one based on a sample of 1,000 individuals and the other on a sample of 5,000—maintaining the original inference structure while enhancing relevance for our target population.

We repeated this process to populate the item pool with examples covering a range of topics, formats, and levels of complexity. Our overarching aim was to ensure comprehensive domain coverage while generating a distribution of item difficulties appropriate for a broad respondent base.

To evaluate the feasibility of deploying the measure in practice, we conducted a small-scale pilot during a multi-day workshop on Data Science and Evidence-Based Policymaking. The pilot sample included professional policymakers, allowing us to observe how respondents interacted with the survey and question formats in a realistic applied setting.

During this pilot, we administered a subset of 11 draft items representing different question types—including true/false, multiple choice, and items with accompanying visuals. Although the pilot did not cover the full item pool, it allowed us to assess survey length, logistical constraints, and respondent behavior. The lessons from this process informed both the wording and format of the final items. Based on timing and usability data, we opted for concise formats that minimized respondent burden while preserving inferential value. We decided to maintain items accompanied by visuals despite slightly longer response times, as they uniquely capture skills in evidence interpretation via graphs, tables, and plots.

Following the pilot, we finalized an initial pool of 52 items to undergo formal evaluation. We designed this item pool with several goals in mind. First, we aimed to capture foundational competencies relevant to real-world (policy) reasoning with data and evidence. This structure enables flexible use of the instrument for researchers and practitioners, such as selecting specific subscales aligned to targeted applications. We will illustrate this with a use case isolating visualization literacy items for tasks involving the prediction and analysis of graphical data.

Second, we constructed items aiming for a range of coverage of difficulty levels, with an intentional tilt toward accessibility. This design choice reflects our expectation that many respondents may have limited formal training in science, statistics, or data analysis. Given the diversity of our target population, which includes the general public and government professionals, we prioritized items that would avoid ceiling or floor effects and allow for discrimination across the entire ability range.

Third, we prioritized "efficient" item formats based on the lessons learned from the pilot. These included fixed-choice formats that were quick to complete, along with selected visual formats that provided richer inference targets despite their greater time demands.

Finally, we deliberately excluded "don't know" or "prefer not to answer" options. While this may introduce response biases—particularly if certain subgroups are more likely (or willing) to guess—it reflects a trade-off intended to encourage engagement and produce full response data. We acknowledge that this choice may have implications for differential item functioning, which we consider in the analyses and interpretation of our results.

A complete list of the 52 draft items is provided in the Supplementary Appendix.

Data collection

We administered the test inventory to respondents recruited through the survey research platform Prolific. Our sampling strategy targeted two broad populations in two countries: general population respondents and government workers from the United States and the United Kingdom. This dual-frame approach allowed us to compare performance across groups with varying levels of policy engagement.

Data collection was conducted in two identical survey waves: the first in July and the second in December 2024. Across both waves, we obtained a total of 470 valid responses, consisting of 338 general population participants and 132 government workers.

Following recruitment, respondents were redirected to our custom-built survey hosted in Qualtrics. To mitigate respondent fatigue while ensuring full item coverage, each participant was randomly assigned a subset of 20 items from the full 52-item test inventory. We stratified item blocks to ensure representation from each of the four subdomains—scientific literacy, statistical literacy, causal reasoning, and data and visualization literacy—within each participant's test. This blocked randomization allowed us to collect a balanced response set across the full item pool.

In addition to responses to the scientific reasoning items, we collected background information on participants' educational attainment, professional domain, and self-reported familiarity with quantitative methods, including math, statistics, and data analysis. These variables served both descriptive and analytical purposes. Table 1 summarizes key demographic and background characteristics across the general and government subsamples.

To validate the scale, we also collected data from two additional samples. The first sample was a group of students and policy professionals ($n = 21$ in total) enrolled in data science courses at a Public Policy school. Students were taking a semester-long introductory course in data science, while professionals participated in a three-day module on Data and AI literacy. As with the Prolific sample, participants in these groups received a subset of 20 items from the full test inventory. We also collected additional self-assessment data relevant for validation, focusing on their capacity to use evidence in policy professional settings.

The second validation sample was collected as part of a pre-election study for the German federal elections in January and February 2025. A total of 18,681 respondents were recruited both via a survey provider and ads on social media (Facebook and Instagram). The study focused on expectations regarding the outcome of the federal election in order to conduct citizen forecasting (Lewis-Beck and Skalaban, 1989). Two items from the inventory were integrated in order to test the predictiveness of data literacy skills for forecasting performance. The results of this validation are also presented below.

To support validation of the inventory, all respondents in the Prolific online sample also completed an applied reasoning task. In this task, participants were shown two brief summaries of hypothetical research studies and asked to select which of the two was better suited to inform a specific policy question. This served as an external benchmark for evaluating the relationship between scientific reasoning skills and policy-relevant decision-making.

Characteristic	UK General Population N = 169	US General Population N = 169	UK Government Workers N = 66	US Government Workers N = 66
Gender				
Female	97 (58%)	90 (54%)	46 (70%)	45 (68%)
Male	69 (42%)	76 (46%)	20 (30%)	21 (32%)
Highest academic degree				
High school/some college or less	1 (0.6%)	5 (3.0%)	1 (1.5%)	1 (1.5%)
Bachelor's degree	104 (62%)	89 (53%)	41 (62%)	30 (45%)
Master's degree	52 (31%)	64 (38%)	23 (35%)	30 (45%)
Doctoral degree	12 (7.1%)	11 (6.5%)	1 (1.5%)	5 (7.6%)
Formal training in math, stats, data analysis				
None	30 (18%)	10 (5.9%)	13 (20%)	8 (12%)
Minimal amount	68 (40%)	71 (42%)	32 (48%)	33 (50%)
Moderate amount	50 (30%)	53 (31%)	19 (29%)	16 (24%)
Significant amount	21 (12%)	35 (21%)	2 (3.0%)	9 (14%)
Field of academic degree				
STEM	44 (26%)	25 (15%)	12 (18%)	7 (11%)
Life & environmental sciences	25 (15%)	55 (33%)	12 (18%)	10 (15%)
Social sciences & education	30 (18%)	41 (24%)	20 (30%)	24 (36%)
Business, law & administration	14 (8.3%)	4 (2.4%)	2 (3.0%)	6 (9.1%)
Humanities & arts	37 (22%)	23 (14%)	11 (17%)	8 (12%)
Other	19 (11%)	21 (12%)	9 (14%)	11 (17%)

Table 1: Sample characteristics. The table summarizes the samples used for item analysis and selected validation analyses. Participants were recruited via Prolific and screened based on (a) country of residence (United Kingdom or United States) and (b) employer type (government vs. non-government).

Item analysis

IRT modeling

To establish the psychometric properties of the inventory, we implemented Item Response Theory (IRT) modeling using dichotomous response data from our UK and US population samples. Given that the scale is designed to measure latent abilities related to individuals' capacity to engage with and interpret scientific evidence, IRT offers clear advantages over Classical Test Theory (CTT), particularly its ability to provide item-level information that is invariant across populations and to model the probability of a correct response as a function of underlying ability.

While most existing studies on scientific reasoning do not employ psychometric models in scale development, those that do tend to rely heavily on Rasch or one-parameter logistic (1PL) models (Edelsbrunner and Dablander, 2019). However, as they note, broader IRT models such as the 2PL and 3PL can offer deeper psychological insight by accounting for variation in item discrimination and guessing behavior—factors that can affect validity by revealing why certain items may underperform or elicit unintended response behaviors.

We estimated and compared three nested IRT models: the one-parameter logistic model (1PL), which assumes equal discrimination across items; the two-parameter logistic model (2PL), which allows item discrimination to vary; and the three-parameter logistic model (3PL), which additionally accounts for guessing behavior.

We implemented the models using marginal maximum likelihood estimation with the `mirt` package in R (Chalmers, 2012). Based on likelihood ratio tests and information criteria (AIC/BIC), we selected the 2PL model providing the best balance of parsimony and explanatory power for our data (see Supplementary Tables B3-B5). Additionally, this model offers a framework for comparing item quality in our specific measurement setup, where the respondents answered only a subset of 20 items (i.e., incomplete-block test design).

Substantively, while the items span distinct literacies—scientific literacy, statistical literacy, data and visualization literacy, and causal reasoning—we conceptualize these as interrelated facets of a broader construct of scientific evidence consumption literacy. Thus, we opted for a uni-dimensional IRT model. Methodologically, the 2PL model offers a balance between model complexity and parameter stability, which is particularly important given our sample size (n=470). Introducing a multidimensional IRT model would substantially increase the number of parameters to estimate, potentially compromising model reliability and interpretability in the absence of strong prior information or substantially larger sample sizes. Considering raw performance scores from our sample that we calculated for both the entire set of items and domain-specific subsets, we find substantial correlations among these sub-domains, providing suggestive evidence that a unidimensional model is appropriate for capturing a general reasoning ability (see Supplementary Table B6). Nonetheless, we acknowledge that multidimensional IRT or bifactor models could offer alternative modeling frameworks that allow for both general and domain-specific latent traits. These could be explored in future work, particularly when larger samples or targeted research questions justify the added complexity.

Item selection

Our goal in the item evaluation phase was to identify a subset of items that effectively discriminate between levels of the ability while maintaining balanced coverage across the four subdomains: causal reasoning, data and visualization, scientific, and statistical literacy. We considered two fit-based criteria for item retention: First, we prioritized items with discrimination (a) parameters greater than 0.7, as these items more effectively distinguish between respondents at different levels of latent ability. Second, to detect potential local dependence, we examined residual correlations using Yen's Q_3 statistic. Specifically, we flagged item pairs within the same subdomain with Q_3 values exceeding 0.2. In such cases, we retained the item with the higher discrimination parameter.

Following our selection criteria, we excluded 13 items for having discrimination parameters below 0.7. These items exhibited relatively flat item characteristic curves, meaning they were less effective at distinguishing between respondents with differing levels of the underlying evidence consumption ability.

An additional nine items were flagged under the local independence criterion. Specifically, we identified item pairs within the same subdomain whose residual correlations exceeded the Q_3 threshold of 0.2, suggesting redundancy or dependence beyond what the IRT model accounts for. In such cases, we retained the item with higher discrimination and removed its lower-performing counterpart. This step helps preserve the unidimensionality assumption of the model and improves the interpretability of scores.

Table 2 provides an overview of the 30 retained items from the test inventory, grouped by domain. For each item, we report: (1) the proportion of respondents who answered the item correctly; (2) the median time taken to respond; (3) the number of respondents who saw the item, given the randomized block design; and (4) the item parameters from a two-parameter logistic (2PL) model, including discrimination (a)—how well the item differentiates between respondents of differing ability)—and difficulty (b) the latent trait level at which the item has a 50% chance of being answered correctly. Supplementary Table B1 presents an analogue table of the excluded items, which indicates the reason for exclusion: either low discrimination (below the 0.7 threshold) or local dependence (i.e., the lower-performing item in a high-Q_3 item pair).

The consolidated inventory offers several desirable properties. Substantively, the selected items cover a wide range of topics, both across and within the four sub-domains. The items vary in difficulty, with more challenging items typically found in the data and visualization literacy sub-domain, while relatively easier items appear more frequently in the scientific literacy sub-domain. In practice, it may be advisable to exclude particularly easy items, as they may generate little information about individuals' abilities. However, this decision should be guided by the expected ability distribution in the target population; for this reason, we did not impose a fixed difficulty threshold during item selection. Another strength of the inventory is its generally low cost in terms of survey time. On average, items took 35 seconds to answer (median: 23 seconds). Items in the data and visualization literacy sub-domain tend to be systematically more time-intensive to field, as they present respondents with information-rich visualizations and scenarios. Whether to include such items is ultimately a practical trade-off: if these tasks are of interest, survey designs should allow for the expected 30–60 seconds of response time per item.

domains most closely aligned with applied scientific reasoning—namely, scientific methods, scientific evidence consumption, and cause-and-effect reasoning. These results offer suggestive evidence of convergent validity, under the logic that participants who see themselves as more knowledgeable in evidence-relevant domains also perform better on the INSPIRE assessment.

Importantly, the modest size of these correlations is consistent with prior work showing that individuals' self-assessments of scientific reasoning skills are often imprecise and influenced by overconfidence (Drummond Otten and Fischhoff, 2023). That said, these associations reinforce the claim that INSPIRE taps into a construct that respondents themselves associate with scientific competence in real-world decision-making.

Next, for the policy student and professional sample undergoing data science training, we examined correlations between θ scores and self-assessments of their ability to perform tasks essential for working with scientific data in policy contexts (bottom panel of Table 3). INSPIRE scores showed moderate to strong positive correlations with all self-assessed abilities. Specifically, the strongest associations were observed for abilities directly related to data interpretation and application: "describe necessary data to answer policy questions", "choose methods to answer policy questions", and "understand technical language in scientific studies". We also find slightly lower, but still meaningful, correlations for the remaining items.

These patterns reinforce the convergent validity of the INSPIRE inventory, indicating that individuals who score higher on the assessment also perceive themselves as more capable in practical, evidence-based policy work, particularly in areas involving data comprehension and application.

Known-groups validity In addition to self-perceptions, we examined differences in θ scores across groups defined by educational attainment and level of empirical training to establish known-groups validity for the Prolific sample. Figure 2 presents the distribution of scores across these groups. As one might expect, participants with higher levels of education demonstrated higher INSPIRE scores. A similar pattern emerges when comparing participants by their reported exposure to empirical methods during their education: those with no or minimal empirical training had lower scores than those reporting moderate or significant training in mathematics, statistics, or data analysis. These patterns are consistent with theoretical expectations and support the scale's known-groups validity: individuals with

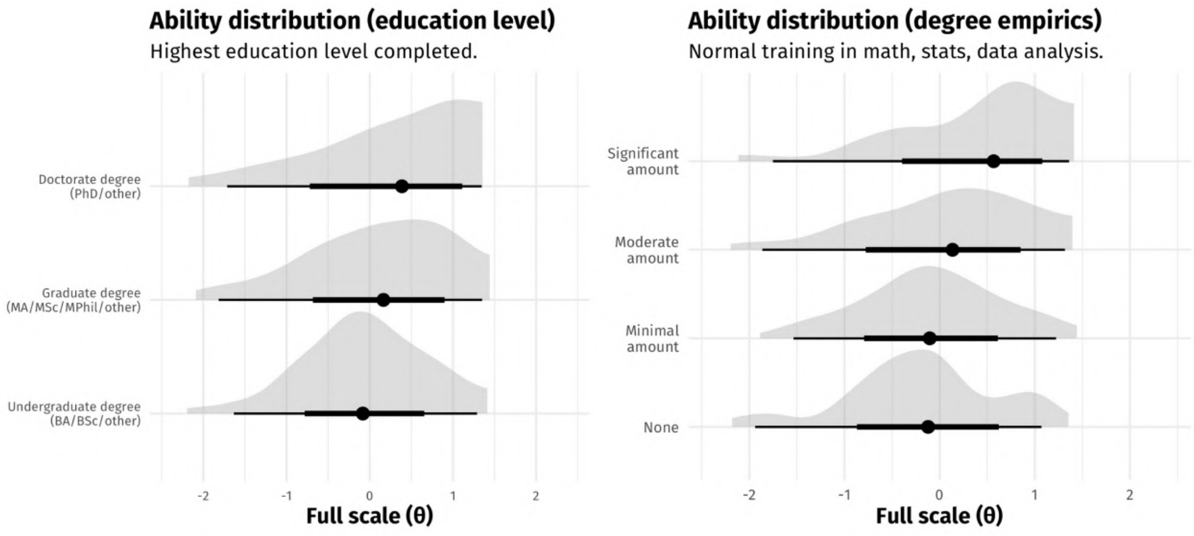


Figure 2: Distribution of ability scores across groups. Density plots of θ score distributions across groups defined by education level (left) and empirical training (right). Lines indicate group means with 66% and 95% confidence intervals.

more extensive academic or applied exposure to empirical methods perform better on the types of reasoning tasks assessed by INSPIRE.

To formally quantify the relationship between these background characteristics and INSPIRE scores, we modeled scores as a linear function of participants' education level, formal training in math, statistics, and data analysis, age, and sex. The results of this ordinary least squares (OLS) regression are presented in Supplementary Table C1. While formal training in empirical methods showed a significant positive association with INSPIRE scores (e.g., individuals with 'significant amount' of training scoring notably higher than those with none), education level, age, and sex did not reach conventional levels of statistical significance after accounting for other factors. Importantly, the overall model explained only a small proportion of the variance in INSPIRE scores ($R^2 = 0.059$). This suggests that while INSPIRE scores are predictably associated with relevant prior experiences (like formal empirical training), they capture a substantial amount of individual variation in scientific evidence literacy that is not explained by general demographic or educational background, reinforcing the unique utility of the instrument.

Together, these findings suggest that the inventory is sensitive to both self-perceived and objectively measured variations in scientific evidence literacy. The inventory appears to capture a domain-specific form of applied reasoning plausibly shaped by training, exposure, and practice in evidence-based environments.

Concurrent validity Next, we examined its association with two distinct external criteria: participants' performance on an applied scientific reasoning task and their ability to reject inaccurate scientific beliefs. These analyses provide evidence for the concurrent validity of the INSPIRE instrument, demonstrating its capacity to predict relevant real-world outcome that require critical engagement with features scientific information.

First, we evaluated how well INSPIRE scores predict performance on an applied reasoning task that required participants to select the better-designed scientific study for a given policy question. Participants were presented with three pairs of hypothetical studies (e.g., comparing a randomized controlled trial to an observational study for assessing an after-school program's impact on academic performance) and asked to identify the one better suited to provide useful evidence. This applied task directly assesses the ability to discern methodological rigor and external validity, skills crucial for evidence-informed policy. We modeled the probability of correctly choosing the optimal study design as a function of INSPIRE θ scores using a mixed-effects logistic regression, with respondent random effects to account for repeated measures.

As shown in Figure 3 (left panel), higher INSPIRE scores were associated with a significantly increased probability of correctly identifying the most suitable study design for the policy question (OR = 2.52, 95% CI=[1.89, 3.37], p=<0.001). To assess the contribution of INSPIRE scores beyond basic demographic characteristics, we also ran a model controlling for education, age, and sex. In this controlled model, the association remained robust (OR = 2.58, 95% CI=[1.93, 3.45], p=<.001), indicating that the inventory captures a distinct ability to critically evaluate research methodology in a policy-relevant context.

Second, we investigated whether INSPIRE scores predict participants' ability to identify and reject inaccurate statements on controversial scientific issues (Drummond and Fischhoff, 2017b). This test was delivered to a subset of Prolific respondents (n=250; implemented in the December 2024 data collection wave). We administered eight statements on topics like climate change, GMOs, and nuclear power, with participants indicating whether each statement was true or false. These statements were designed such that the scientifically accurate answer was unambiguous (e.g., "Human activities, such as burning fossil fuels, are the primary cause of recent global climate change"). We then modeled the probability of rejecting an inaccurate scientific belief as a function of INSPIRE θ scores, again using a mixed-effects logistic regression with respondent random effects. The results, presented in Figure 3 (right panel), demonstrate a positive relationship: participants with higher INSPIRE scores were more likely to identify and reject inaccurate scientific claims

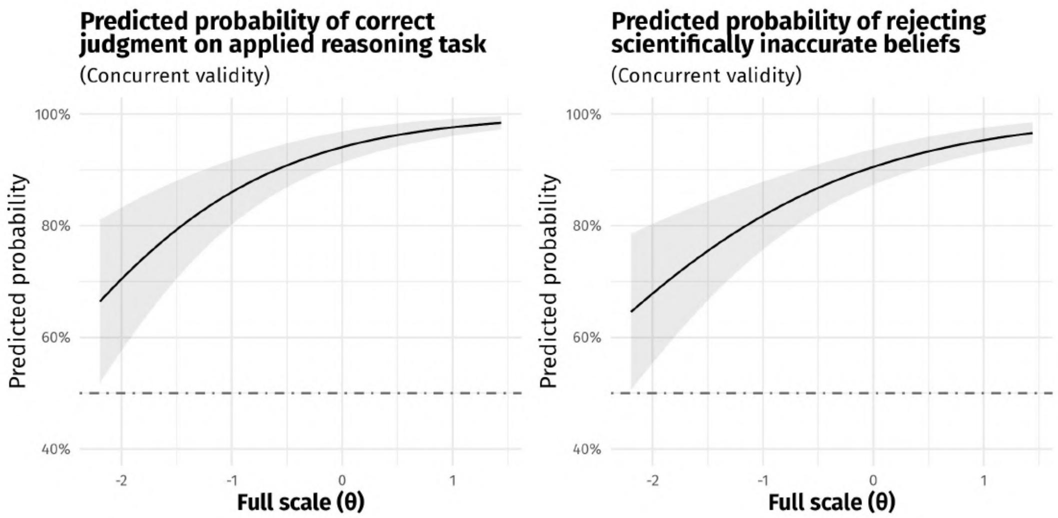


Figure 3: Predicted probabilities of INSPIRE scores on applied reasoning task and scientific belief accuracy. Results from multilevel logistic regression models with respondent random effects estimating the relationship between participants' INSPIRE θ scores and outside criteria.

correctly. This relationship also held when controlling for education, age, and sex (OR = 2.13, 95% CI=[1.64, 2.75], p=<.001), suggesting that the INSPIRE inventory captures an individual's propensity to align with established scientific consensus, further supporting its concurrent validity. The full model details are provided in Supplementary Tables C2 and C3.

Predictive validity Beyond concurrent associations, we also assessed the inventory's predictive utility in a real-world forecasting context. This analysis employed a separate sample from a pre-election study conducted in the context of the 2025 German federal elections (for more details, see Rajski et al., 2025). This study specifically focused on citizen forecasting, integrating three items from the INSPIRE inventory to test the predictiveness of data and visual literacy skills for forecasting performance.

For this validation, we use data from roughly 15,500 participants, recruited from a commercial online access panel as well as via ads run on Facebook and Instagram in January and February 2025, responded to two INSPIRE visual literacy items: one requiring the identification of the correct visual representation of exponentially growing raw population data (DVL04 - Visual data representation 1), and another asking them to identify a key insight that can be learn from a data chart (DVL05 - Visual information 1). In addition, they responded to an item that was later screened out in our item selection procedure (DVL09 -Complementary probability 3), but that still provides a useful indicator of their basic understanding of probability, very similar to INSPIRE items DVL01 and DVL02.

Through a multi-stage development process, we constructed an initial pool of 52 items tailored to these demands, focusing on empirical, quantitative evidence and key aspects of scientific literacy, statistical literacy, data and visualization literacy, and causal reasoning. Using Item Response Theory (IRT), we identified a final 30-item inventory that demonstrates strong psychometric properties, including high discrimination and coverage across a meaningful range of difficulties. The inventory is particularly effective at precisely distinguishing individuals at lower-to-average levels of scientific evidence consumption literacy, a characteristic valuable for identifying those most in need of interventions or for assessing baseline competencies in broad populations.

While not proposing an entirely new concept, INSPIRE systematically measures a set of integrated reasoning abilities that, as our validation exercises demonstrate, relate to, but are distinct from, general knowledge or demographic profiles. Our comprehensive validation across three distinct samples provided compelling evidence for the inventory's utility. Specifically, a) INSPIRE scores showed expected relationships with relevant external criteria; and b) despite expected associations with prior experiences, demographic, and educational factors, INSPIRE demonstrates significant unique and incremental explanatory power of critical applied outcomes.

Drawing from the extensive analyses presented in this study, we argue that INSPIRE inventory provides a psychometrically sound, contextually relevant, and practically scalable tool for measuring scientific evidence consumption literacy. It captures nuanced abilities that extend beyond general knowledge or commonly collected demographic profiles, offering a valuable instrument for both research and applied settings by providing measurement where existing scales may fall short.

Guidance for implementation of INSPIRE

The results of our item analysis and validation exercises provide the basis for the use of the INSPIRE inventory. However, administering the full instrument is likely infeasible with samples of professionals, and something we would advise against in most real-world applications. Instead, researchers can operate with subsets of the scale. Even small item sets can be informative, as illustrated in the pre-election forecast validation exercise.

When customizing the scale, several criteria should be considered. First is content. Different items tap into different skills and dimensions of scientific evidence consumption literacy. Consequently, their predictive validity and causal relationships with external variables are likely to vary depending on the underlying skill. Second, resource constraints play a role. Some items, particularly those providing high information (often including visual components), are more time-intensive for respondents. If efficiency is a priority, shorter items may be preferable.

Third, item difficulty should be taken into account. The ability distribution in the target population may differ from that of our validation sample. Our item analysis suggests that the inventory performs particularly well in the lower-ability range, which may be precisely where measurement precision is needed. However, in contexts where higher levels of scientific literacy are expected—such as among students in higher education—researchers may consider administering other items of the full 52-item inventory. Some of the items excluded from the validated core set still offer acceptable psychometric properties and provide greater difficulty. Fourth, item discrimination should guide selection. Items with higher discriminatory power are preferable and should be prioritized. Finally, where feasible, the reliability of any customized version of the scale should be assessed prior to deployment, for example, through a pilot study.

Researchers might also consider implementing computerized adaptive testing to optimally select items from the inventory as a function of participants' expected abilities (Montgomery and Rossiter, 2022).

Another practical question concerns that of scoring. In principle, researchers could use raw scores–i.e., the proportion of correctly solved items–as a proxy for respondents' evidence consumption literacy. This assumes equal weighting across items and ignores information about difficulty and discrimination, but can still work reasonably well, in particular with a balanced set of items. Alternatively, ability estimates can be derived from an IRT model applied to the new sample. In our Prolific sample, raw scores were highly correlated with ability estimates ($r = 0.96$).

Finally, the INSPIRE items were designed to make them agnostic towards policy or political contexts. Researchers might want to translate the INSPIRE scale where needed, however, we recommend re-evaluating the psychometric properties of any translated version to ensure validity and reliability in the target language and context.

Applications for INSPIRE in research and practice

The INSPIRE scale opens the door to a range of applications that go beyond measurement. As a descriptive tool, it can illuminate how these skills are distributed across political, demographic, and professional lines. Prior research shows that scientific literacy and trust in science vary along partisan lines, particularly in polarized contexts such as climate change and public health crises (Funk et al., 2019; Cologna et al., 2025; Funk et al., 2019; Cologna et al., 2025). INSPIRE allows for a more granular investigation of whether these divides also exist in the ability to process scientific evidence—both among policymakers and in the general population. Similarly, the scale can shed light on how competencies differ across professional backgrounds or educational levels, adding empirical depth to the nexus between expertise and decision-making in public policy (Christensen, 2021; Cantarelli, Belle, and Hall, 2023; Christensen, 2021; Cantarelli, Belle, and Hall, 2023).

Beyond mapping differences, the scale lends itself to explanatory research into how evidence influences broader political and policy-related outcomes. For instance, greater capacity to understand and critically engage with scientific evidence may foster trust in science itself, increase demand for high-quality evidence, and shape concrete policy preferences and decisions downstream (Hjort et al., 2021). INSPIRE offers a way to test these mechanisms directly and to examine whether evidence consumption literacy can predict better policy outcomes across sectors.

INSPIRE can also be treated as an outcome in its own right. Scholars have tested different interventions to understand what can boost policymakers' ability to make sense of research evidence (Crowley et al., 2021; Mehmood, Naseer, and Chen, 2024; Crowley et al., 2021; Mehmood, Naseer, and Chen, 2024). INSPIRE provides a metric to evaluate the effectiveness of such interventions more rigorously.

Finally, INSPIRE also holds potential for applied use beyond academia. In professional policymaking environments—whether in government ministries, legislative offices, or think tanks—it could function as a monitoring and evaluation instrument. Institutions could use it to assess baseline competencies in evidence use, target professional development, or inform recruitment and promotion processes. Just as literacy and numeracy assessments have become commonplace in workforce development, evidence consumption skills may emerge as a core competency in modern governance. With its flexibility and domain-specific grounding, INSPIRE stands to contribute to a more evidence-informed policymaking culture.

Acknowledgments

The authors acknowledge funds by a grant from the Hertie School Faculty Activity Fund. SRR acknowledges financial support by Hertie School in the form of a doctoral stipend.

References

Inhelder, Bärbel and Jean Piaget (1958). *The growth of logical thinking: From childhood to adolescence*. Ed. by Anne Parsons and Stanley Milgram. The growth of logical thinking: From childhood to adolescence. New York, NY, US: Basic Books.

Miller, Jon D. (1983). “Scientific Literacy: A Conceptual and Empirical Review”. *Daedalus* 112.2, pp. 29–48.

Lewis-Beck, Michael S and Andrew Skalaban (1989). “Citizen forecasting: can voters see into the future?” *British Journal of Political Science* 19.1, pp. 146–153.

Wallman, Katherine K. (1993). “Enhancing Statistical Literacy: Enriching Our Society”. *Journal of the American Statistical Association* 88.421, pp. 1–8.

National Academies of Sciences, Engineering, and Medicine (1996). *National Science Education Standards*. Washington, D.C.: National Academies Press.

Watson, Jane and Rosemary Callingham (2003). “STATISTICAL LITERACY: A COMPLEX HIERARCHICAL CONSTRUCT”. *STATISTICS EDUCATION RESEARCH JOURNAL* 2.2, pp. 3–46.

Economic Co-operation, Organisation for and Development (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. OECD Publishing.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.

Schield, Milo (2010). “Assessing Statistical Literacy: Take CARE”. In: *Assessment Methods in Statistical Education*. John Wiley & Sons, Ltd, pp. 133–152.

DeBoer, George E. (2011). “The globalization of science education”. *Journal of Research in Science Teaching* 48.6, pp. 567–591.

Chalmers, R. Philip (2012). “mirt: A Multidimensional Item Response Theory Package for the R Environment”. *Journal of Statistical Software* 48.6, pp. 1–29.

Gormally, Cara, Peggy Brickman, and Mary Lutz (2012). “Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates’ Evaluation of Scientific Information and Arguments”. *CBE—Life Sciences Education* 11.4, pp. 364–377.

Boy, Jeremy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete (2014). “A Principled Way of Assessing Visualization Literacy”. *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 1963–1972.

Fischer, Frank, Ingo Kollar, Stefan Ufer, Beate Sodian, Heinrich Hussmann, Reinhard Pekrun, Birgit Neuhaus, Birgit Dorner, Sabine Pankofer, Martin Fischer, Jan-Willem Strijbos, Moritz Heene, and Julia Eberle (2014). “Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education”. *Frontline Learning Research* 2.3, pp. 28–45.

Juncker, Jean-Claude (2014). *Mission letter to Commissioner Moedas*.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh (2015). “Knowledge does not protect against illusory truth”. *Journal of Experimental Psychology. General* 144.5, pp. 993–1002.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

D’Ignazio, Catherine and Rahul Bhargava (2016). “DataBasic: Design Principles, Tools and Activities for Data Literacy Learners”. *The Journal of Community Informatics* 12.3.

Wolff, Annika, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem (2016). “Creating an Understanding of Data Literacy for a Data-driven Society”. *The Journal of Community Informatics* 12.3.

Drummond, Caitlin and Baruch Fischhoff (2017a). “Development and Validation of the Scientific Reasoning Scale”. *Journal of Behavioral Decision Making* 30.1, pp. 26–38.

— (2017b). “Individuals with greater science literacy and education have more polarized beliefs on controversial science topics”. *Proceedings of the National Academy of Sciences* 114.36, pp. 9587–9592.

Lee, Sukwon, Sung-Hee Kim, and Bum Chul Kwon (2017). “VLAT: Development of a Visualization Literacy Assessment Test”. *IEEE Transactions on Visualization and Computer Graphics* 23.1, pp. 551–560.

Opitz, Ansgar, Moritz Heene, and Frank Fischer (2017). “Measuring scientific reasoning – a review of test instruments”. *Educational Research and Evaluation* 23.3-4, pp. 78–101.

Romine, William L., Troy D. Sadler, and Andrew T. Kinslow (2017). “Assessment of scientific literacy: Development and validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR)”. *Journal of Research in Science Teaching* 54.2, pp. 274–295.

Waldmann, Michael (2017). *The Oxford Handbook of Causal Reasoning*. Oxford University Press.

Pearl, Judea and Dana Mackenzie (2018). *The book of why: the new science of cause and effect*. New York: Basic Books.

Sabbag, Anelise, Joan Garfield, and Andrew Zieffler (2018). "ASSESSING STATISTICAL LITERACY AND STATISTICAL REASONING: THE REALI INSTRUMENT". *STATISTICS EDUCATION RESEARCH JOURNAL* 17.2, pp. 141–160.

Börner, Katy, Andreas Bueckle, and Michael Ginda (2019). "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments". *Proceedings of the National Academy of Sciences* 116.6, pp. 1857–1864.

Edelsbrunner, Peter A. and Fabian Dablander (2019). "The Psychometric Modeling of Scientific Reasoning: a Review and Recommendations for Future Avenues". *Educational Psychology Review* 31.1, pp. 1–34.

Funk, Cary, Meg Hefferon, Brian Kennedy, and Courtney Johnson (2019). *Partisanship influences views on the role and value of scientific experts in policy debates*.

U.S. Congress (2019). *H.R. 4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act*.

Christensen, Johan (2021). "Expert knowledge and policymaking: a multi-disciplinary research agenda". *Policy & Politics* 49.3, pp. 455–471.

Crowley, D. Max, J. Taylor Scott, Elizabeth C. Long, Lawrie Green, Azaliah Israel, Lauren Supplee, Elizabeth Jordan, Kathryn Oliver, Shannon Guillot-Wright, Brittany Gay, Rachel Storace, Naomi Torres-Mackie, Yolanda Murphy, Sandra Donnay, Jenna Reardanz, Rebecca Smith, Kristina McGuire, Elizabeth Baker, Ana Antonopoulos, Mary McCauley, and Cagla Giray (2021). "Lawmakers' use of scientific evidence can be improved". *Proceedings of the National Academy of Sciences* 118.9, e2012955118.

European Commission (2021). *European citizens' knowledge and attitudes towards science and technology: report.* LU: Publications Office.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini (2021). "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities". *American Economic Review* 111.5, pp. 1442–1480.

Linden, Sander van der (2022). "Misinformation: susceptibility, spread, and interventions to immunize the public". *Nature Medicine* 28.3, pp. 460–467.

Montgomery, Jacob M. and Erin L. Rossiter (2022). *Adaptive Inventories: A Practical Guide for Applied Researchers*. 1st ed. Cambridge University Press.

Cantarelli, Paola, Nicola Belle, and Jeremy L. Hall (2023). "Information use in public administration and policy decision-making: A research synthesis". *Public Administration Review* 83.6, pp. 1667–1686.

Drummond Otten, Caitlin and Baruch Fischhoff (2023). "Calibration of scientific reasoning ability". *Journal of Behavioral Decision Making* 36.3, e2306.

Ge, Lily W., Yuan Cui, and Matthew Kay (2023). "CALVI: Critical Thinking Assessment for Literacy in Visualizations". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Hamburg Germany: ACM, pp. 1–18.

Hornberger, Marie, Arne Bewersdorff, and Claudia Nerdel (2023). "What do university students know about Artificial Intelligence? Development and validation of an AI literacy test". *Computers and Education: Artificial Intelligence* 5, p. 100165.

Vivalt, Eva and Aidan Coville (2023). "How do policymakers update their beliefs?" *Journal of Development Economics* 165, p. 103121.

Wang, Bingcheng, Pei-Luen Patrick Rau, and Tianyi Yuan (2023). "Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale". *Behaviour & Information Technology* 42.9, pp. 1324–1337.

Weber, Patrick, Marc Pinski, and Lorenz Baum (2023). *Toward an Objective Measurement of AI Literacy*.

Mehmood, Sultan, Shaheen Naseer, and Daniel L Chen (2024). "Training Policymakers in Econometrics".

Ng, Davy Tsz Kit, Wenjie Wu, Jac Ka Lok Leung, Thomas Kin Fung Chiu, and Samuel Kai Wah Chu (2024). "Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach". *British Journal of Educational Technology* 55.3, pp. 1082–1104.

Toma, Mattie and Elizabeth Bell (2024). "Understanding and increasing policymakers' sensitivity to program impact". *Journal of Public Economics* 234, p. 105096.

Tyson, Alec and Brian Kennedy (2024). *Public Trust in Scientists and Views on Their Role in Policymaking*. Tech. rep. Pew Research Center.

Cologna, Viktoria et al. (2025). "Trust in scientists and their role in society across 68 countries". *Nature Human Behaviour*, pp. 1–18.

Rajski, Hannah, Cornelius Erfort, Thomas Gschwend, Elias Koch, Simon Munzert, and Lukas Stoetzer (2025). *Pre-election survey "Citizen Forecasting of the German Federal Election 2025"*. https://zweitstimme.org.

Ramirez-Ruiz, Sebastian and Roman Senninger (2025). *Policy documents across 185 countries predominantly rely on evidence from the Global North*.

Rimkutė, Dovilė and Honorata Mazepus (2025). "Citizens' perceptions of the legitimacy of independent agencies: The effects of expertise-based and reputation-sourced authority". *Public Administration Review*.

Inhelder, Bärbel and Jean Piaget (1958). *The growth of logical thinking: From childhood to adolescence*. Ed. by Anne Parsons and Stanley Milgram. The growth of logical thinking: From childhood to adolescence. New York, NY, US: Basic Books.

Miller, Jon D. (1983). "Scientific Literacy: A Conceptual and Empirical Review". *Daedalus* 112.2, pp. 29–48.

Lewis-Beck, Michael S and Andrew Skalaban (1989). "Citizen forecasting: can voters see into the future?" *British Journal of Political Science* 19.1, pp. 146–153.

Wallman, Katherine K. (1993). "Enhancing Statistical Literacy: Enriching Our Society". *Journal of the American Statistical Association* 88.421, pp. 1–8.

National Academies of Sciences, Engineering, and Medicine (1996). *National Science Education Standards*. Washington, D.C.: National Academies Press.

Watson, Jane and Rosemary Callingham (2003). "STATISTICAL LITERACY: A COMPLEX HIERARCHICAL CONSTRUCT". *STATISTICS EDUCATION RESEARCH JOURNAL* 2.2, pp. 3–46.

Economic Co-operation, Organisation for and Development (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. OECD Publishing.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Schield, Milo (2010). "Assessing Statistical Literacy: Take CARE". In: *Assessment Methods in Statistical Education*. John Wiley & Sons, Ltd, pp. 133–152.

DeBoer, George E. (2011). "The globalization of science education". *Journal of Research in Science Teaching* 48.6, pp. 567–591.

Chalmers, R. Philip (2012). "mirt: A Multidimensional Item Response Theory Package for the R Environment". *Journal of Statistical Software* 48.6, pp. 1–29.

Gormally, Cara, Peggy Brickman, and Mary Lutz (2012). "Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments". *CBE—Life Sciences Education* 11.4, pp. 364–377.

Boy, Jeremy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete (2014). "A Principled Way of Assessing Visualization Literacy". *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 1963–1972.

Fischer, Frank, Ingo Kollar, Stefan Ufer, Beate Sodian, Heinrich Hussmann, Reinhard Pekrun, Birgit Neuhaus, Birgit Dorner, Sabine Pankofer, Martin Fischer, Jan-Willem Strijbos, Moritz Heene, and Julia Eberle (2014). "Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education". *Frontline Learning Research* 2.3, pp. 28–45.

Juncker, Jean-Claude (2014). *Mission letter to Commissioner Moedas*.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh (2015). "Knowledge does not protect against illusory truth". *Journal of Experimental Psychology. General* 144.5, pp. 993–1002.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

D'Ignazio, Catherine and Rahul Bhargava (2016). "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners". *The Journal of Community Informatics* 12.3.

Wolff, Annika, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem (2016). "Creating an Understanding of Data Literacy for a Data-driven Society". *The Journal of Community Informatics* 12.3.

Drummond, Caitlin and Baruch Fischhoff (2017a). "Development and Validation of the Scientific Reasoning Scale". *Journal of Behavioral Decision Making* 30.1, pp. 26–38.

Drummond, Caitlin and Baruch Fischhoff (2017b). "Individuals with greater science literacy and education have more polarized beliefs on controversial science topics". *Proceedings of the National Academy of Sciences* 114.36, pp. 9587–9592.
Lee, Sukwon, Sung-Hee Kim, and Bum Chul Kwon (2017). "VLAT: Development of a Visualization Literacy Assessment Test". *IEEE Transactions on Visualization and Computer Graphics* 23.1, pp. 551–560.
Opitz, Ansgar, Moritz Heene, and Frank Fischer (2017). "Measuring scientific reasoning – a review of test instruments". *Educational Research and Evaluation* 23.3-4, pp. 78–101.
Romine, William L., Troy D. Sadler, and Andrew T. Kinslow (2017). "Assessment of scientific literacy: Development and validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR)". *Journal of Research in Science Teaching* 54.2, pp. 274–295.
Waldmann, Michael (2017). *The Oxford Handbook of Causal Reasoning*. Oxford University Press.
Pearl, Judea and Dana Mackenzie (2018). *The book of why: the new science of cause and effect*. New York: Basic Books.
Sabbag, Anelise, Joan Garfield, and Andrew Zieffler (2018). "ASSESSING STATISTICAL LITERACY AND STATISTICAL REASONING: THE REALI INSTRUMENT". *STATISTICS EDUCATION RESEARCH JOURNAL* 17.2, pp. 141–160.
Börner, Katy, Andreas Bueckle, and Michael Ginda (2019). "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments". *Proceedings of the National Academy of Sciences* 116.6, pp. 1857–1864.
Edelsbrunner, Peter A. and Fabian Dablander (2019). "The Psychometric Modeling of Scientific Reasoning: a Review and Recommendations for Future Avenues". *Educational Psychology Review* 31.1, pp. 1–34.
Funk, Cary, Meg Hefferon, Brian Kennedy, and Courtney Johnson (2019). *Partisanship influences views on the role and value of scientific experts in policy debates*.
U.S. Congress (2019). *H.R. 4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act*.
Christensen, Johan (2021). "Expert knowledge and policymaking: a multi-disciplinary research agenda". *Policy & Politics* 49.3, pp. 455–471.
Crowley, D. Max, J. Taylor Scott, Elizabeth C. Long, Lawrie Green, Azaliah Israel, Lauren Supplee, Elizabeth Jordan, Kathryn Oliver, Shannon Guillot-Wright, Brittany Gay, Rachel Storace, Naomi Torres-Mackie, Yolanda Murphy, Sandra Donnay, Jenna Reardanz, Rebecca Smith, Kristina McGuire, Elizabeth Baker, Ana Antonopoulos, Mary McCauley, and Cagla Giray (2021). "Lawmakers' use of scientific evidence can be improved". *Proceedings of the National Academy of Sciences* 118.9, e2012955118.
European Commission (2021). *European citizens' knowledge and attitudes towards science and technology: report.* LU: Publications Office.
Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini (2021). "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities". *American Economic Review* 111.5, pp. 1442–1480.
Linden, Sander van der (2022). "Misinformation: susceptibility, spread, and interventions to immunize the public". *Nature Medicine* 28.3, pp. 460–467.
Montgomery, Jacob M. and Erin L. Rossiter (2022). *Adaptive Inventories: A Practical Guide for Applied Researchers*. 1st ed. Cambridge University Press.
Cantarelli, Paola, Nicola Belle, and Jeremy L. Hall (2023). "Information use in public administration and policy decision-making: A research synthesis". *Public Administration Review* 83.6, pp. 1667–1686.
Drummond Otten, Caitlin and Baruch Fischhoff (2023). "Calibration of scientific reasoning ability". *Journal of Behavioral Decision Making* 36.3, e2306.
Ge, Lily W., Yuan Cui, and Matthew Kay (2023). "CALVI: Critical Thinking Assessment for Literacy in Visualizations". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Hamburg Germany: ACM, pp. 1–18.
Hornberger, Marie, Arne Bewersdorff, and Claudia Nerdel (2023). "What do university students know about Artificial Intelligence? Development and validation of an AI literacy test". *Computers and Education: Artificial Intelligence* 5, p. 100165.
Vivalt, Eva and Aidan Coville (2023). "How do policymakers update their beliefs?" *Journal of Development Economics* 165, p. 103121.

Wang, Bingcheng, Pei-Luen Patrick Rau, and Tianyi Yuan (2023). “Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale”. *Behaviour & Information Technology* 42.9, pp. 1324–1337.

Weber, Patrick, Marc Pinski, and Lorenz Baum (2023). *Toward an Objective Measurement of AI Literacy*.

Mehmood, Sultan, Shaheen Naseer, and Daniel L Chen (2024). “Training Policymakers in Econometrics”.

Ng, Davy Tsz Kit, Wenjie Wu, Jac Ka Lok Leung, Thomas Kin Fung Chiu, and Samuel Kai Wah Chu (2024). “Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach”. *British Journal of Educational Technology* 55.3, pp. 1082–1104.

Toma, Mattie and Elizabeth Bell (2024). “Understanding and increasing policymakers’ sensitivity to program impact”. *Journal of Public Economics* 234, p. 105096.

Tyson, Alec and Brian Kennedy (2024). *Public Trust in Scientists and Views on Their Role in Policymaking*. Tech. rep. Pew Research Center.

Cologna, Viktoria et al. (2025). “Trust in scientists and their role in society across 68 countries”. *Nature Human Behaviour*, pp. 1–18.

Rajski, Hannah, Cornelius Erfort, Thomas Gschwend, Elias Koch, Simon Munzert, and Lukas Stoetzer (2025). *Pre-election survey "Citizen Forecasting of the German Federal Election 2025"*. https://zweitstimme.org.

Ramirez-Ruiz, Sebastian and Roman Senninger (2025). *Policy documents across 185 countries predominantly rely on evidence from the Global North*.

Rimkutė, Dovilė and Honorata Mazepus (2025). “Citizens’ perceptions of the legitimacy of independent agencies: The effects of expertise-based and reputation-sourced authority”. *Public Administration Review*.

Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory

Online Appendix

Contents

Appendix A The INSPIRE inventory

A.1 Validated 30-item bank

Scientific Literacy (8 items)

SEL1. Scientific Hypothesis
Which of the following best describes a scientific hypothesis?
A. A proven fact
B. A detailed description of an experiment
C. **A statement that can be tested ✓**
D. A summary of research findings

SEL2. Scientific Consensus
Scientific consensus means that there is unanimous agreement among scientists on a particular issue.
- True
- **False ✓**

SEL3. Quality of Evidence
*Which of the following criteria is usually **NOT** relevant for assessing the quality of research evidence for policy decision-making?*
A. The relevance of the studies to the specific policy question
B. The consistency of findings across different studies
C. **The number of mentions of individual studies by news media ✓**
D. The strength of study designs

SEL4. Research Credibility
Which of the following would raise concerns about the credibility of a research article?
A. The article presents findings that are consistent with studies from other contexts
B. **The article does not provide detailed information about its research methodology and data sources ✓**
C. The article is freely accessible to the public
D. The article uses a large sample size in its research

SEL5. Scientific Practice 1
Do you think that the following scenario represents appropriate scientific practice?

A group of scientists is asked to review grant proposals. They base their funding recommendations on the researcher's experience, the soundness of the research design plans, and preliminary data from the research proposals submitted.
- **This represents appropriate scientific practice ✓**
- This does not represent appropriate scientific practice

SEL6. Scientific Practice 2
Do you think that the following scenario represents appropriate scientific practice?

A group of scientists is selected to help conduct a state-funded research study on monetary policy. The scientists and the state agency providing the funding agree in advance that results of the study will be published if they align with the government's stance on the policy.
- This represents appropriate scientific practice
- **This does not represent appropriate scientific practice ✓**

SEL7. Trustworthy Data
When evaluating the reliability of a data source, which of the following factors is the least important?
A. The methodology used to collect the data
B. The sample size of the data
C. The date when the data was collected
D. **The font used in the data presentation ✓**

SEL8. Trustworthy Sources
The government has announced the plan to implement a certain policy. The policy may have significant side effects, so you do some research to determine the effectiveness of the policy option compared to alternatives.
Which of the following sources would likely provide the most accurate information?

A. The website of an interest group
B. An opinion piece about the policy solution on the news
C. **A research study conducted by independent researchers ✓**
D. Information from a trusted friend from a country that implemented the policy solution six months ago

Statistical Literacy (7 items)

SML1. Center and Spread
Which of the following statements IS NOT true for the given data?
2, 2, 6, 8, 12

A. The mean (average) is 6
B. **The median is 3 ✓**
C. The mode is 2
D. All the statements are correct

SML2. Sample Sizes and Uncertainty
Two studies estimate the average educational level of two rural areas. Each study uses the survey on a random sample of inhabitants. Study 1 uses 1,000 respondents, and study 2 uses 5,000 respondents. Which statement is true?

A. The estimate of the average educational level from each study will be equally uncertain
B. The uncertainty in the estimate of the average educational level will be smaller in study 1 than in study 2
C. **The uncertainty in the estimate of the average educational level will be larger in study 1 than in study 2 ✓**
D. None of the above

SML3. Base Rate 1
A test shows promising results in providing early detection for infections of a new flu variant. However, 5% of all test results are falsely positive; that is, results indicate that infection is present when the patient is, in fact, infection-free. Given this false positive rate, how many people out of 10,000, none of whom carries the new flu variant, would have a false positive result and be alarmed unnecessarily?

A. 5
B. 35
C. 50
D. **500 ✓**

SML4. ML Accuracy
The accuracy of a machine learning model is solely dependent on the complexity of the algorithm used, not the quality of the training data.

A. **False ✓**
B. True

SML5. ML and Perpetuated Bias
If a machine learning model is trained on biased data, the model's predictions are likely to perpetuate those biases.

A. **True ✓**
B. False

SML6. ML and Flawed Predictions
A public policy researcher is using a machine learning model to predict the success of social programs. They notice that the model systematically generates wrong predictions for a recent migrant group. What could be the most likely reason?

A. **The training data does not generalize onto the group of migrants ✓**
B. The algorithm used is outdated
C. The model's performance is being evaluated incorrectly
D. The training data was too large to process effectively

SML7. ML Application
Which of the following is a common application of machine learning?
A. Making phone calls
B. Browsing the internet
C. **Predicting future trends based on past data ✓**
D. Charging electronic devices

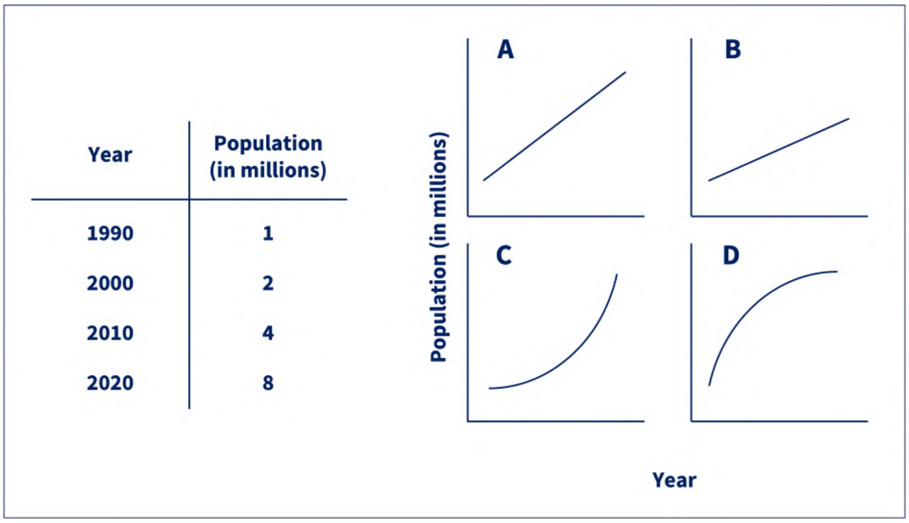
Data and Visualization Literacy (8 items)

DVL1. Complementary Probability 1
If the probability of an event happening is 0.2, what is the probability of the event NOT happening?
A. 0
B. 0.5
C. **0.8 ✓**
D. That is impossible to tell

DVL2. Complementary Probability 2
In a city, 60% of the students participate in an after-school program. Assuming all residents are either students or non-students and that one resident is selected at random, what is the probability that the selected resident does not participate in an after-school program?
A. 0.2
B. **0.4 ✓**
C. 0.6
D. That is impossible to tell

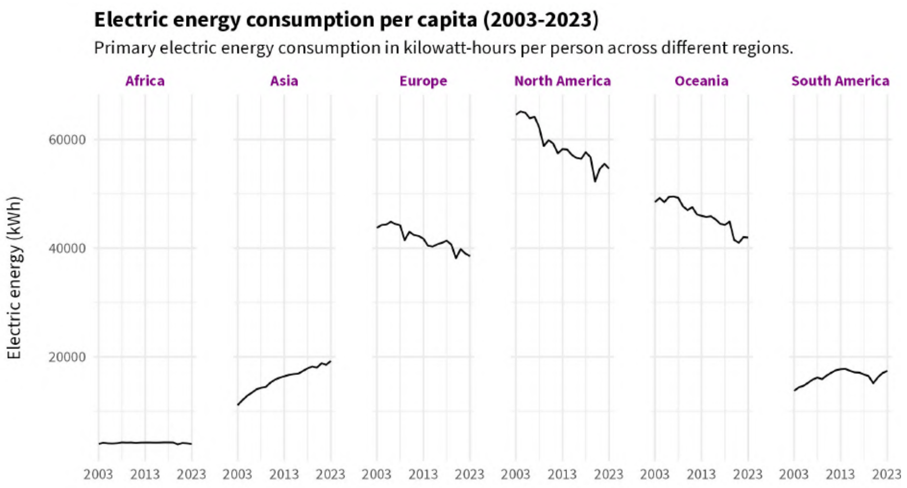
DVL3. Percentages
An earthquake destroyed 40% of the houses in a coastal city. After assessments by the construction authorities, 20% of the remaining houses were deemed inhabitable due to structural damage. What percentage of the original houses is still habitable after these two events?
A. 40%
B. **48% ✓**
C. 60%
D. This cannot be calculated without knowing the original number of houses

DVL4. Visual Data Representation 1
You have a table tracking population over time. Which graph best represents the data?



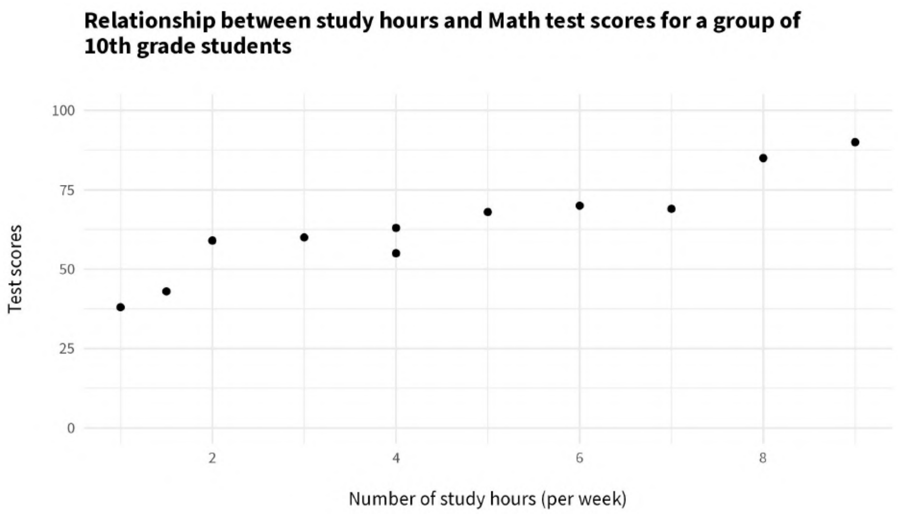
A. A
B. B
C. **C ✓**
D. D

DVL5. Visual Information 1
Examine the graph. What can be learned?



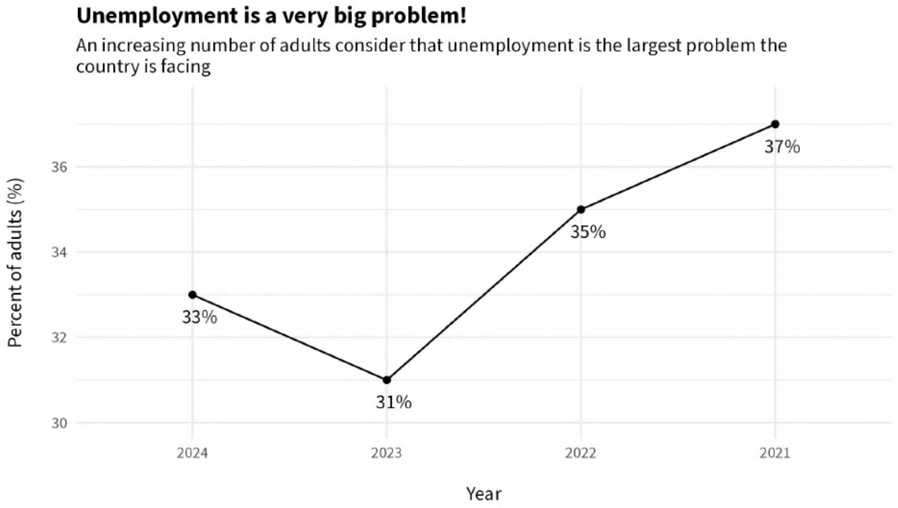
A. A global decrease in electric energy consumption
B. Asia will surpass Europe in per capita use by 2050
C. **North America is the highest per capita electric consumer ✓**
D. Renewables caused a decrease in consumption in Europe, North America, and Oceania

DVL6. Visual Information 2
Examine the graph. What can be learned?



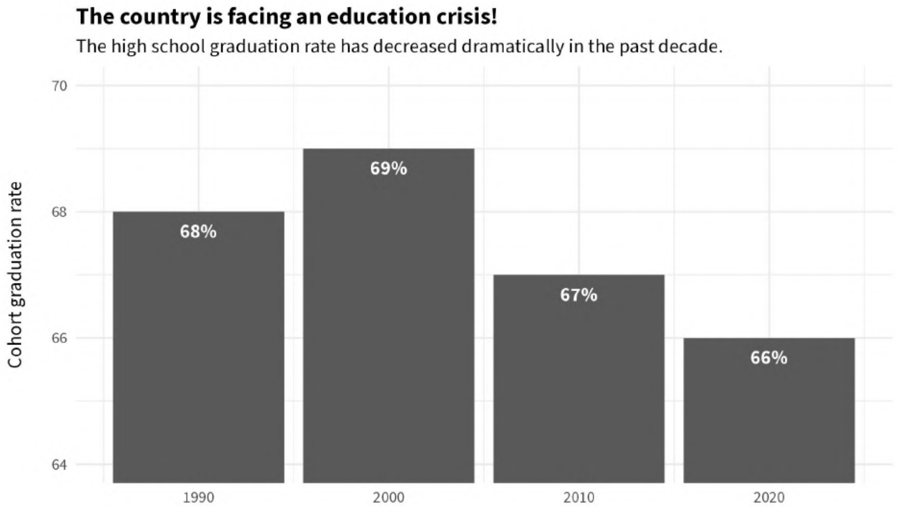
A. **Positive relationship between study hours and test scores ✓**
B. Negative relationship
C. No relationship
D. Cannot be determined

DVL7. Flawed Visuals 1
Examine the graph carefully. Which IS NOT a problematic feature?



A. The x-axis is inverted
B. **Data points shown with low precision ✓**
C. Axis does not start at 0%
D. Subtitle is inconsistent with trend

DVL8. Flawed Visuals 2
Examine the graph carefully. Which IS NOT a problematic feature?



A. Unclear year summaries in bars
B. Axis does not start at 0
C. **Subtitle exaggerates a marginal change ✓**
D. Graph lacks color, reducing clarity

Causal Reasoning (7 items)

CR1. Correlation vs. Causation
When two variables are correlated, it means that changes in one variable are directly causing changes in the other variable.
A. True
B. **False ✓**

CR2. Correlation
Which of the following statements best describes statistical correlation?
A. Correlation is a measure of the cause-and-effect relationship between two variables.
B. Correlation measures the difference between two groups.
C. Correlation measures the impact of one variable on another over time.
D. **Correlation measures the strength and direction of the relationship between two variables. ✓**

CR3. Learning from Correlation
A study found a correlation coefficient of 0.8 between a person's years of education and their income level. What can be concluded from the study?
A. Higher income causes more years of education.
B. **There is a strong statistical relationship between years of education and income level. ✓**
C. More years of education directly lead to higher income.
D. There is a strong causal relationship between years of education and income level.

CR4. Randomized Control Trials
In randomized control trials (RCTs), participants are typically randomly assigned to either the treatment or the control group.
A. **True ✓**
B. False

CR5. Confounder
Consider the statement: "Survey data revealed that people who sleep with their shoes on are much more likely to wake up with a headache." What is problematic about this conclusion?
A. There might be reverse causality
B. The sample size is too small
C. There is no control group
D. **Something else might cause people both to sleep with their shoes on and to wake up with a headache ✓**

CR6. Causal Effect of a Policy
Which option best describes a situation where a policy has a causal effect on unemployment?
A. Unemployment fell after the policy was implemented
B. The policy is correlated with changes in the unemployment rate
C. **The policy directly causes changes in the unemployment rate, beyond what would have happened without the policy ✓**
D. The policy was implemented at the same time as changes in the unemployment rate

CR7. Study Conclusions 1
Does the evidence support the conclusion below?
Evidence: A study found that areas with more green spaces have lower reported stress. Parks are also linked to higher household incomes.
Conclusion: Lower stress causes higher household income.
A. **The conclusion is not supported by the evidence ✓**
B. The conclusion is supported by the evidence

A.2 Item performance

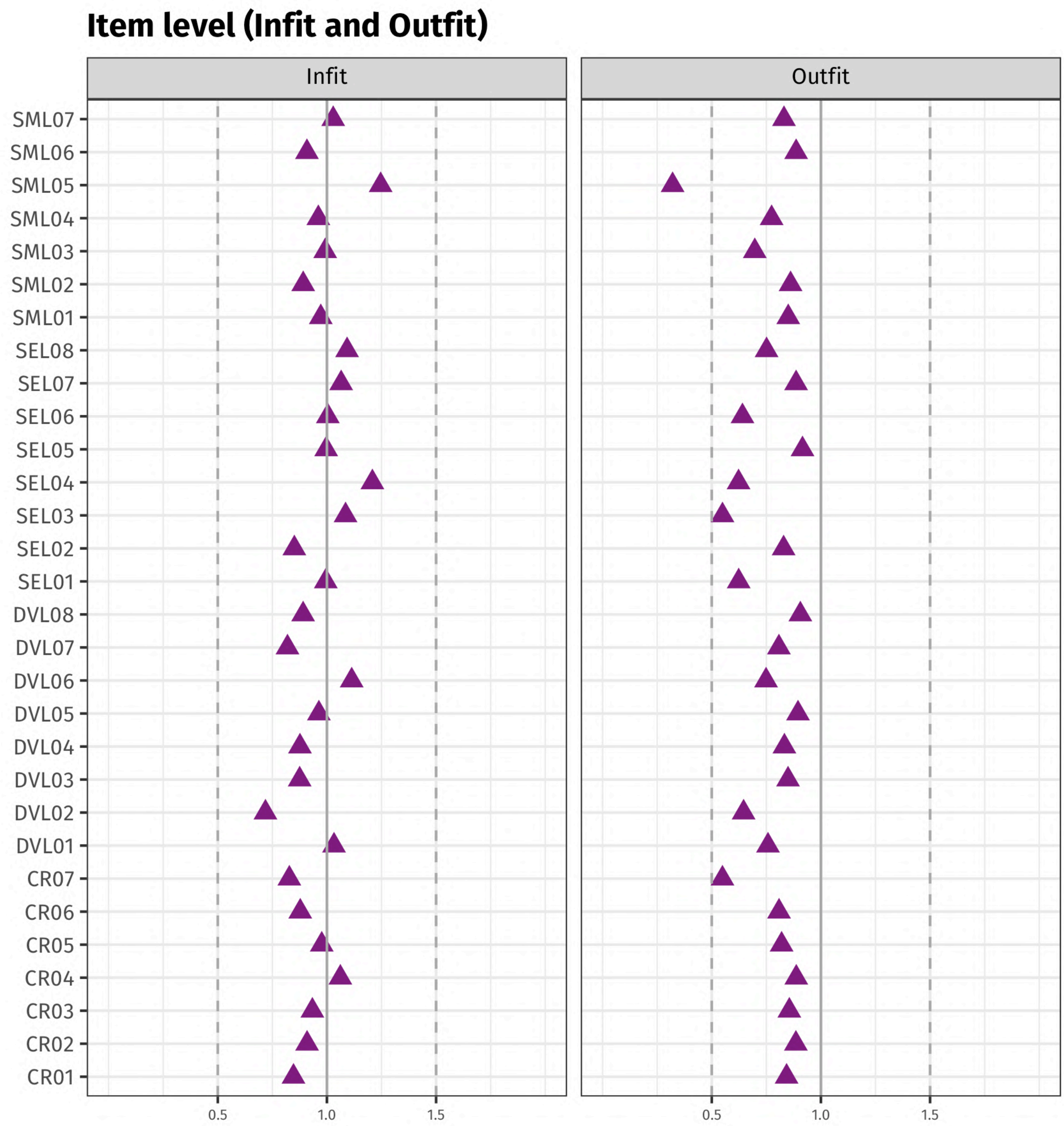


Figure A1: Item Fit Statistics (Infit and Outfit Mean Square). This figure displays the Infit and Outfit Mean Square (MNSQ) values for each of the 30 items in the INSPIRE inventory. The solid vertical line at 1.0 represents ideal item fit, while the dashed vertical lines at 0.5 and 1.5 indicate the conventional range for productive measurement. Values within this range suggest that items are neither underfitting nor overfitting the IRT model, supporting the psychometric quality of the INSPIRE inventory.

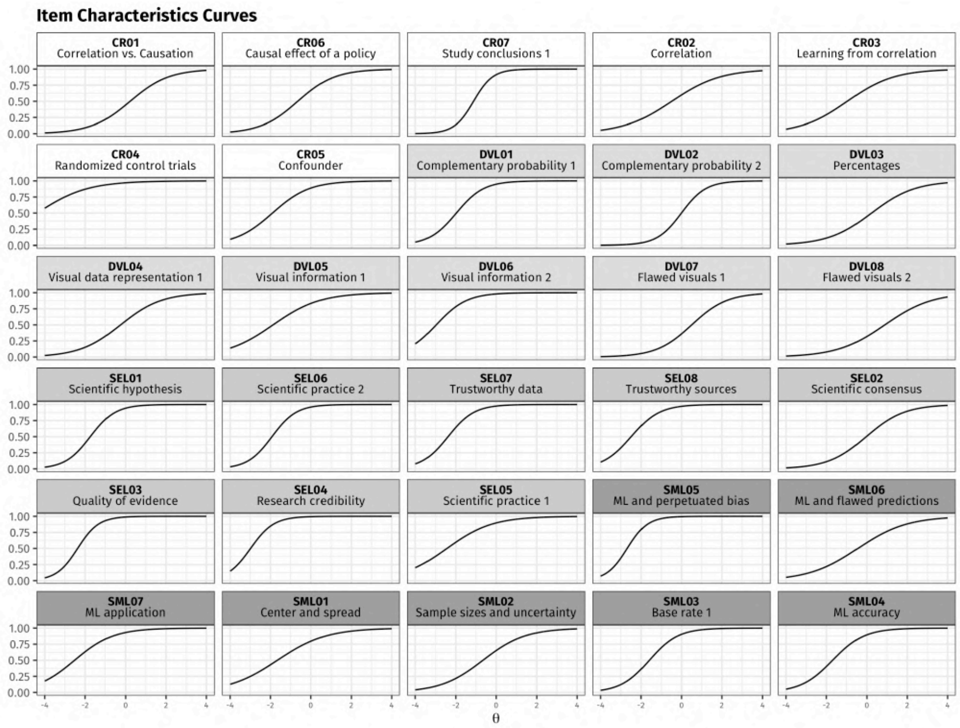


Figure A2: Disaggregated Item Characteristic Curves (ICCs) for each INSPIRE item. This figure presents the individual Item Characteristic Curves (ICCs) for items retained in the INSPIRE inventory. Each curve illustrates the probability of a correct response as a function of the latent ability (θ) score. The varying slopes and positions of these curves demonstrate the items' capacity to discriminate between individuals at different ability levels.

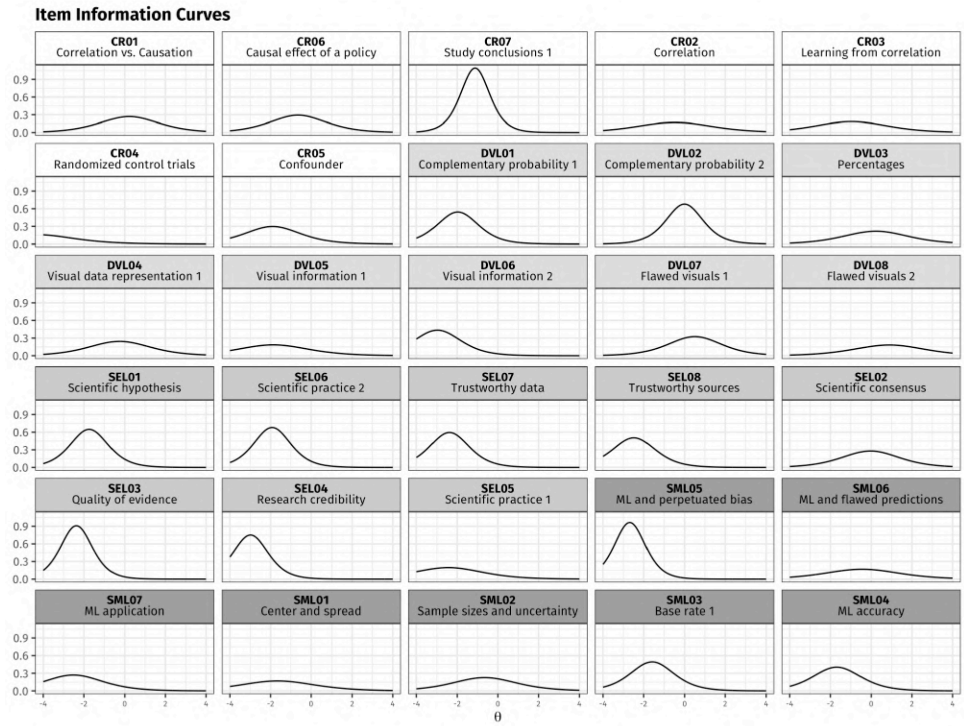


Figure A3: Disaggregated Item Information Curves (IICs) for each INSPIRE item. This figure presents the individual Item Information Curves (IICs) for items retained in the INSPIRE inventory. Each curve illustrates the amount of information an item contributes across the latent ability (θ). The peaks of these curves generally correspond to the item's difficulty, indicating the ability level at which the item provides the most measurement precision.

A.3 Excluded items

Scientific Literacy (7 items)

SEL9 (e). Peer Review Process
In scientific practice, peer review is a process where other experts in the field evaluate the quality and validity of a scientist's research before it is published.
A. **True ✓**
B. False

SEL10 (e). Replicability
Which of the following is a characteristic of a well-designed scientific research study?
A. It always confirms the hypothesis
B. It includes as many variables as possible
C. It relies solely on qualitative data
D. **It is repeatable and can be replicated by others ✓**

SEL11 (e). Peer Review Bias
Peer review ensures that scientific research is free from bias and errors.
A. True
B. **False ✓**

SEL12 (e). Evidence Synthesis
Evidence synthesis involves combining findings from multiple studies to reach conclusions.
A. **True ✓**
B. False

SEL13 (e). Scientific Practice 3
Do you think that the following scenario represents appropriate scientific practice?

Scenario: The national wildlife authority reviews its list of protected and endangered species in response to new academic research findings.
A. **This represents appropriate scientific practice ✓**
B. This does not represent appropriate scientific practice

SEL14 (e). Scientific Practice 4
Do you think that the following scenario represents appropriate scientific practice?

Scenario: The legislature stops funding a widely used sex-education program after several studies show limited effectiveness of the program.
A. **This represents appropriate scientific practice ✓**
B. This does not represent appropriate scientific practice

SEL15 (e). Trustworthy Science
An important factor influencing the way you would categorize a research article as trustworthy science is:
A. The presence of data or graphs
B. **The article was evaluated by unbiased third-party experts (peer review) ✓**
C. The host institution of the researchers
D. The media attention of the study

Statistical Literacy (8 items)

SML8 (e). Mean Calculation
What is the mean (average) of the following set of numbers?
2, 4, 6, 8, 10
A. 4
B. **5 ✓**
C. 6
D. 7

SML9 (e). Median Calculation
What is the median value of the following set of numbers?
4, 7, 2, 6, 12, 8, 10
A. **6 ✓**
B. 7
C. 8
D. 12

SML10 (e). Base Rate 2

A city uses a surveillance AI tool to detect stolen bikes. Theft is very rare, less than 1 in 100,000 bikes get stolen. The AI tool correctly identifies a stolen bike 99% of the time and incorrectly flags a non-stolen bike as stolen 1% of the time.

If the system flags 100 bikes as suspicious, roughly how many of these are likely to be actual stolen bikes?

A. **Less than 5 ✓**
B. Between 5 and 20
C. 50
D. 99

SML11 (e). Base Rate 3

A city has 1,000 inhabitants, 950 of whom are non-smokers and 50 are smokers. A medical test correctly identifies smokers 90% of the time and non-smokers 80% of the time.

If a randomly selected person tests positive for smoking, is that person more likely to be a smoker or a non-smoker?

A. Smoker
B. **Non-smoker ✓**
C. Either option is equally likely
D. The question cannot be answered with the information at hand

SML12 (e). ML Definition

What is machine learning?

A. A way for humans to learn from machines
B. **A method for computers to learn from data ✓**
C. A technique to clean data
D. A type of machine maintenance

SML13 (e). ML and Overfitting

In machine learning, overfitting occurs when a prediction model performs well on training data but poorly on new, unseen data.

A. **True ✓**
B. False

SML14 (e). ML and Supervised Learning

What is the primary goal of supervised learning in machine learning?

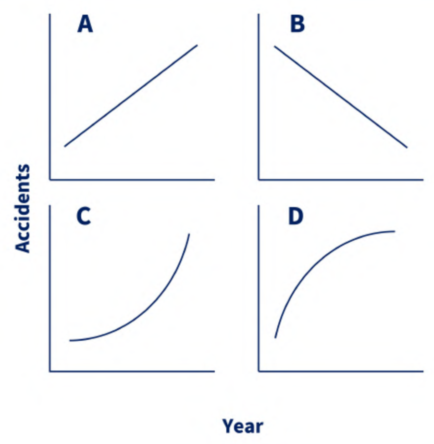
A. To find patterns in data without any labels
B. To prevent humans from being harmed by machine learning
C. **To learn from labeled data to make predictions ✓**
D. To reduce the dimensionality of the data

Data and Visualization Literacy (2 items)

DVL10 (e). Visual Data Representation 2

You have been keeping track of the number of car accidents over time in your city. The table below shows your rough count of these events. Which graph provides the most suitable representation of your data?

Year	Accidents
1990	400
2000	300
2010	200
2020	100



A. A
B. **B ✓**
C. C
D. D

	Election poll interpretation (binary)			
	Pooled item score (standardized)	Visual data representation 1 (DVL04)	Visual information 1 (DVL05)	Complementary probability 3 (DVL09e)
Pooled item score	0.11***			
	[0.11, 0.12]			
DVL04		0.11***		
		[0.10, 0.13]		
DVL05			0.14***	
			[0.12, 0.15]	
DVL09 (e)				0.17***
				[0.15, 0.18]
Educ: Medium	0.05***	0.07***	0.06***	0.06***
	[0.02, 0.07]	[0.04, 0.10]	[0.03, 0.09]	[0.03, 0.08]
Educ: High	0.10***	0.15***	0.14***	0.14***
	[0.07, 0.13]	[0.13, 0.18]	[0.11, 0.17]	[0.11, 0.17]
Gender: Female	-0.03	-0.05	-0.07	-0.06
	[-0.16, 0.11]	[-0.19, 0.08]	[-0.21, 0.06]	[-0.19, 0.08]
Gender: Male	0.02	0.01	0.00	0.00
	[-0.11, 0.16]	[-0.13, 0.14]	[-0.14, 0.13]	[-0.13, 0.14]
Age: Middle	0.04***	0.02+	0.02+	0.02*
	[0.02, 0.07]	[0.00, 0.05]	[0.00, 0.05]	[0.00, 0.05]
Age: Old	0.07***	0.03*	0.04**	0.04**
	[0.05, 0.10]	[0.00, 0.06]	[0.01, 0.06]	[0.01, 0.06]
Political interest	0.03***	0.03***	0.03***	0.03***
	[0.03, 0.04]	[0.03, 0.04]	[0.03, 0.04]	[0.03, 0.04]
Perceived party performance	-0.03***	-0.04***	-0.04***	-0.04***
	[-0.03, -0.03]	[-0.04, -0.04]	[-0.04, -0.03]	[-0.04, -0.03]
Num.Obs.	15570	15570	15570	15570
R2	0.182	0.152	0.157	0.165
R2 Adj.	0.181	0.151	0.156	0.165
RMSE	0.45	0.46	0.45	0.45

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

Table C4: Predicted performance on election poll interpretation.

	First vote prediction accuracy (standardized)			
	Pooled item score (standardized)	Visual data representation 1 (DVL04)	Visual information 1 (DVL05)	Complementary probability 3 (DVL09e)
Pooled item score	0.09***			
	[0.07, 0.10]			
DVL04		0.06***		
		[0.03, 0.09]		
DVL05			0.11***	
			[0.09, 0.14]	
DVL09 (e)				0.14***
				[0.11, 0.17]
Educ: Medium	0.08**	0.09***	0.09**	0.09**
	[0.03, 0.13]	[0.04, 0.15]	[0.03, 0.14]	[0.03, 0.14]
Educ: High	0.05*	0.10***	0.08**	0.08**
	[0.00, 0.10]	[0.04, 0.15]	[0.03, 0.13]	[0.03, 0.13]
Gender: Female	0.12	0.10	0.09	0.10
	[-0.12, 0.37]	[-0.15, 0.34]	[-0.16, 0.33]	[-0.15, 0.35]
Gender: Male	0.10	0.08	0.08	0.08
	[-0.15, 0.34]	[-0.17, 0.33]	[-0.17, 0.32]	[-0.16, 0.33]
Age: Middle	0.08***	0.07**	0.07**	0.07**
	[0.04, 0.13]	[0.02, 0.11]	[0.02, 0.11]	[0.03, 0.11]
Age: Old	0.16***	0.12***	0.13***	0.13***
	[0.11, 0.20]	[0.07, 0.17]	[0.08, 0.18]	[0.08, 0.18]
Political interest	0.02*	0.02*	0.02**	0.02*
	[0.00, 0.03]	[0.00, 0.03]	[0.00, 0.03]	[0.00, 0.03]
Perceived party performance	-0.18***	-0.18***	-0.18***	-0.18***
	[-0.18, -0.17]	[-0.19, -0.18]	[-0.18, -0.17]	[-0.18, -0.17]
Num.Obs.	15302	15302	15302	15302
R2	0.317	0.312	0.314	0.316
R2 Adj.	0.317	0.312	0.314	0.315
RMSE	0.83	0.83	0.83	0.83

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

Table C5: Predicted performance on first vote prediction.

	Second vote prediction accuracy (standardized)			
	Pooled item score (standardized)	Visual data representation 1 (DVL04)	Visual information 1 (DVL05)	Complementary probability 3 (DVL09e)
Pooled item score	0.05***			
	[0.04, 0.06]			
DVL04		0.03**		
		[0.01, 0.06]		
DVL05			0.06***	
			[0.04, 0.08]	
DVL09 (e)				0.10***
				[0.08, 0.12]
Educ: Medium	0.10***	0.11***	0.11***	0.10***
	[0.06, 0.14]	[0.07, 0.15]	[0.07, 0.15]	[0.06, 0.14]
Educ: High	0.13***	0.16***	0.15***	0.15***
	[0.09, 0.17]	[0.12, 0.20]	[0.11, 0.19]	[0.11, 0.19]
Gender: Female	-0.17+	-0.19*	-0.19*	-0.18+
	[-0.36, 0.02]	[-0.38, 0.00]	[-0.38, 0.00]	[-0.37, 0.01]
Gender: Male	-0.14	-0.15	-0.15	-0.14
	[-0.32, 0.05]	[-0.33, 0.04]	[-0.33, 0.04]	[-0.33, 0.04]
Age: Middle	-0.01	-0.02	-0.02	-0.02
	[-0.04, 0.02]	[-0.06, 0.01]	[-0.05, 0.01]	[-0.05, 0.02]
Age: Old	0.00	-0.03	-0.02	-0.01
	[-0.04, 0.03]	[-0.06, 0.01]	[-0.06, 0.02]	[-0.05, 0.02]
Political interest	0.01*	0.01*	0.01*	0.01*
	[0.00, 0.02]	[0.00, 0.02]	[0.00, 0.02]	[0.00, 0.02]
Perceived party performance	-0.25***	-0.26***	-0.26***	-0.26***
	[-0.26, -0.25]	[-0.26, -0.25]	[-0.26, -0.25]	[-0.26, -0.25]
Num.Obs.	15353	15353	15353	15353
R2	0.606	0.604	0.605	0.606
R2 Adj.	0.606	0.604	0.605	0.606
RMSE	0.63	0.63	0.63	0.63

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

Table C6: Predicted performance on first vote prediction.

	Election events prediction accuracy (standardized)			
	Pooled item score (standardized)	Visual data representation 1 (DVL04)	Visual information 1 (DVL05)	Complementary probability 3 (DVL09e)
Pooled item score	0.09***			
	[0.07, 0.11]			
DVL04		0.10***		
		[0.07, 0.13]		
DVL05			0.13***	
			[0.09, 0.16]	
DVL09 (e)				0.11***
				[0.08, 0.14]
Educ: Medium	0.08**	0.10**	0.09**	0.09**
	[0.02, 0.14]	[0.04, 0.16]	[0.03, 0.15]	[0.03, 0.15]
Educ: High	0.08**	0.12***	0.10***	0.11***
	[0.02, 0.14]	[0.06, 0.18]	[0.05, 0.16]	[0.06, 0.17]
Gender: Female	-0.11	-0.14	-0.15	-0.15
	[-0.40, 0.17]	[-0.42, 0.15]	[-0.44, 0.13]	[-0.43, 0.14]
Gender: Male	0.01	-0.01	-0.01	-0.01
	[-0.28, 0.29]	[-0.29, 0.28]	[-0.30, 0.27]	[-0.30, 0.27]
Age: Middle	0.03	0.01	0.01	0.01
	[-0.02, 0.08]	[-0.04, 0.06]	[-0.04, 0.06]	[-0.04, 0.06]
Age: Old	-0.02	-0.05+	-0.04	-0.05+
	[-0.08, 0.04]	[-0.11, 0.01]	[-0.10, 0.01]	[-0.11, 0.01]
Political interest	0.08***	0.08***	0.08***	0.08***
	[0.07, 0.10]	[0.07, 0.10]	[0.07, 0.10]	[0.07, 0.10]
Perceived party performance	-0.06***	-0.06***	-0.06***	-0.06***
	[-0.06, -0.05]	[-0.07, -0.06]	[-0.07, -0.06]	[-0.07, -0.06]
Num.Obs.	15518	15518	15518	15518
R2	0.090	0.086	0.087	0.086
R2 Adj.	0.090	0.085	0.086	0.086
RMSE	0.96	0.96	0.96	0.96

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

Table C7: Predicted performance on election events prediction accuracy.

Appendix D Software statement

We used R version 4.4.2 (R Core Team, 2024) and the following R packages:

`broom` v. 1.0.7 (Robinson, Hayes, and Couch, 2024)
`dplyr` v. 1.1.4 (Wickham, François, et al., 2023)
`ggdist` v. 3.3.2 (Kay, 2024)
`ggh4x` v. 0.3.0 (van den Brand, 2024)
`ggmirt` v. 0.1.0 (Masur, 2025)
`ggplot2` v. 3.5.1.9000 (Wickham, 2016)
`ggtext` v. 0.1.2 (Wilke and Wiernik, 2022)
`gt` v. 0.11.1 (Iannone et al., 2024)
`gtsummary` (Sjoberg et al., 2021)
`haven` v. 2.5.4 (Wickham, Miller, and Smith, 2023)
`janitor` v. 2.2.0 (Firke, 2023)
`mirt` (Chalmers, 2012)
`readxl` v. 1.4.3 (Wickham and Bryan, 2023)
`stringr` v. 1.5.1 (Wickham, 2023)
`tidyr` v. 1.3.1 (Wickham, Vaughan, and Girlich, 2024)

References

Chalmers, R. Philip (2012). “mirt: A Multidimensional Item Response Theory Package for the R Environment”. *Journal of Statistical Software* 48.6, pp. 1–29.
Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange (2021). “Reproducible Summary Tables with the gtsummary Package”. *The R Journal* 13 (1), pp. 570–580.
Wilke, Claus O. and Brenton M. Wiernik (2022). *ggtext: Improved Text Rendering Support for 'ggplot2'*.
Firke, Sam (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*.
Wickham, Hadley (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*.
Wickham, Hadley and Jennifer Bryan (2023). *readxl: Read Excel Files*.
Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan (2023). *dplyr: A Grammar of Data Manipulation*.
Wickham, Hadley, Evan Miller, and Danny Smith (2023). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*.
Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy (2024). *gt: Easily Create Presentation-Ready Display Tables*.
Kay, Matthew (2024). “ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics”. *IEEE Transactions on Visualization and Computer Graphics* 30.1, pp. 414–424.
R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
Robinson, David, Alex Hayes, and Simon Couch (2024). *broom: Convert Statistical Objects into Tidy Tibbles*.
van den Brand, Teun (2024). *ggh4x: Hacks for 'ggplot2'*.
Wickham, Hadley, Davis Vaughan, and Maximilian Girlich (2024). *tidyr: Tidy Messy Data*.
Masur, Philipp K. (2025). *ggmirt: Plotting functions to extend "mirt" for IRT analyses*.
Chalmers, R. Philip (2012). “mirt: A Multidimensional Item Response Theory Package for the R Environment”. *Journal of Statistical Software* 48.6, pp. 1–29.
Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange (2021). “Reproducible Summary Tables with the gtsummary Package”. *The R Journal* 13 (1), pp. 570–580.
Wilke, Claus O. and Brenton M. Wiernik (2022). *ggtext: Improved Text Rendering Support for 'ggplot2'*.
Firke, Sam (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*.
Wickham, Hadley (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*.
Wickham, Hadley and Jennifer Bryan (2023). *readxl: Read Excel Files*.
Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan (2023). *dplyr: A Grammar of Data Manipulation*.
Wickham, Hadley, Evan Miller, and Danny Smith (2023). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*.
Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy (2024). *gt: Easily Create Presentation-Ready Display Tables*.

Kay, Matthew (2024). “ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics”. *IEEE Transactions on Visualization and Computer Graphics* 30.1, pp. 414–424.
R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
Robinson, David, Alex Hayes, and Simon Couch (2024). *broom: Convert Statistical Objects into Tidy Tibbles*.
van den Brand, Teun (2024). *ggh4x: Hacks for 'ggplot2'*.
Wickham, Hadley, Davis Vaughan, and Maximilian Girlich (2024). *tidyr: Tidy Messy Data*.
Masur, Philipp K. (2025). *ggmirt: Plotting functions to extend "mirt" for IRT analyses*.

Annex. List of papers

- *Ramirez-Ruiz, Sebastian (2025). *Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience.* No. wqbe4. Center for Open Science. (under review)
- *Ramirez-Ruiz, Sebastian and Roman Senninger (2025). *Policy documents across 185 countries predominantly rely on evidence from the Global North.* No. w8q3y. Center for Open Science. (under review)
- Ramirez-Ruiz, Sebastian. *The Bundestag Expert Witness Tracker (BEWIT): A database of German Bundestag public expert hearings.*
- Ramirez-Ruiz, Sebastian and Simon Munzert. *Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory.*

* These manuscripts have been deposited as pre-prints on the Open Science Framework (OSF) repository.