Finding Nemo, Abroad*

Vastav
LSE

June, 2026

Abstract

Scientific research is disproportionately focused on high-income settings. What determines the geography of scientific attention? Incentives within academia—such as publication, citation, and funding structures—as well as the backgrounds and interests of researchers themselves may shape where science is directed. This motivates a central question: should policy aim to attract diverse scientists, or to reward diverse science?

I study whether the internationalization of higher education shifts the direction of scientific research. I exploit the staggered introduction of English-taught master's programs in science departments at the departments of biological sciences at 118 leading universities in non-Anglophone countries between 1990 and 2019. These programs lower language barriers and expand access to a broader pool of international students and early-career researchers. I measure research direction using a novel index that links the species studied in each publication to their native habitats.

I find that program adoption increases research on linguistically dissimilar countries by about 19 additional papers per university per year (more than doubling of research output about these countries), with effects growing over time. Treated universities also expand the set of countries they study (by 2 additional countries a year). These results show that changes in researcher composition can shift the direction of scientific attention. Additional analysis focusing on mechanisms shall provide insight into how supply- and demand-side forces jointly shape the allocation of research effort.

*Acknowledgements: This paper is being written to meet the requirements for EC599 MRes research Paper. Use of Generative AI tools like ChatGPT and Claude are acknowledged.

1 Introduction

Despite addressing global phenomena, modern science remains disproportionately focused on WEIRD settings (Western, Educated, Industrialized, Rich, Democratic). Across a wide range of domains, from agriculture (Suri and Udry (2022); Moscona and Sastry (2025)) and tropical diseases (Association et al. (2004); Dimitris et al. (2021)) to social and economic outcomes (Henrich et al. (2010); Das et al. (2013)), scientific knowledge about developing countries remains strikingly limited, with potentially important consequences for their ability to confront pressing development challenges (De Souza (2022); Acemoglu and Zilibotti (2001); Moscona and Sastry (2025)). The geographical distribution of scientific attention remains resistant to reallocation. On the demand side, the incentive structure of science – encompassing citations, grants, and publications in elite journals – disproportionately rewards research on high-income settings.[1] This makes globally representative science individually costly to pursue. On the supply side, academia draws from a narrow socioeconomic and geographic pool, and researchers tend to work on contexts they know and privilege the societal concerns they have navigated.[2]

A growing class of policies attempts to diversify supply of scientists. For example, scholarships targeting students from developing countries, sex or race-based mentoring programs, scientific diplomacy initiatives, and affirmative action in admissions all aim to change the composition of research workforce. While their proponents place some intrinsic value on diversity, these interventions are justified on the instrumental grounds that changing who does science will change what science gets done. However, researchers who enter through diversity-oriented pipelines face the same publication incentives and funding structures as everyone else. Whether changes in the composition of the scientific workforce can meaningfully alter the direction of scientific attention despite these shared incentives is ultimately an empirical question, with direct consequences for how policymakers design and evaluate initiatives aimed at making science more globally representative.

This paper provides evidence that it can. I study this using the staggered introduction of English-Taught Master's Programs (ETMPs) in 20 non-Anglophone high-income countries (see Figure 4). Before these programs existed, universities drew international students from linguistically connected populations. Students from Francophone Africa, for example, could access French universities, while those from India, Ghana, or Indonesia faced an insuperable language barrier restricting entry of scientific talent. English-medium instruction could lower these barriers, broadening the pool of potential migrants and al-

[1] Research on incentives in science and their effects on topic-choice is well-studied, for example, in Das et al. (2013); Fry and MacGarvie (2024); Azoulay et al. (2019); Hill and Stein (2025a,b).

[2] Descriptive on diversity of scientific and inventor workforce is well-documented in the literature. See Bell et al. (2019); Stansbury and Schultz (2023); Bayer and Rouse (2016); Lundberg and Stearns (2019); Cook (2020). homophily of identity and research topics is documented in Koning et al. (2021); Einiö et al. (2019); Keding and Ritterrath (2025); Dossi (2024).

tering the geographical composition of students entering elite research programs. I use the timing of these program launches as the source of variation in the supply of internationally diverse early-career researchers, and ask whether this compositional shift in the workforce leaves a detectable footprint in the geography of scientific output.

To trace these shifts, I focus on the departments of biological and environmental sciences, where research topics can be credibly linked to geographic contexts through the plants and animals being studied and the locations used to source research samples. This setting is well-suited to the question for three reasons. First, conservation biology has a documented diversity problem in its training pipelines.[3] Second, conducting field research in this domain requires localized skills and ecological familiarity that vary meaningfully across researcher backgrounds. Finally, the researchers operate under a skewed incentive structure prioritizing attention towards charismatic species and accessible field sites in wealthy countries, closely mirroring other disciplines like public health.

I hand-collect the years of launch of hundreds of ETMPs at biology departments of 118 research-intensive universities across 20 Asian and European countries. My study period begins in 1990, when only four of these universities offered an ETMP. By 2010, this had risen to 70 universities, and by 2019, all but seven universities had at least one ETMP (see Figure 3). While this expansion is rapid, it remains modest relative to the growth of English-taught programs in fields such as computer science and business administration. Each treated university is matched to a control university located in one of the "Big Four" destination countries – the United States, United Kingdom, Canada, and Australia – which have historically attracted the largest international student populations.

Using staggered difference-in-differences analysis following Callaway and Sant'Anna (2021), I find that introducing ETMPs significantly diversifies the national composition of incoming early-career researchers (ECRs). In the decade following a program's launch, departments attract ECRs from roughly 2 additional countries per year, with effects peaking at 3 countries by the tenth year – an 18% increase over a baseline of 11 nationalities. The share of incoming ECRs from linguistically dissimilar countries (LD authors) rises by 5 percentage points, from a baseline of 24.5%. Critically, these effects are concentrated among early-career researchers and are absent for mid- and late-career researchers, which helps rule out confounding explanations such as broader departmental internationalization initiatives that might combine ETMPs with diversified senior hiring or expanded international research portfolios.

The effects on research output are equally pronounced. Departments with ETMPs study roughly 2.3 additional countries and expand their taxonomic coverage by 14 plant and 20

[3]For example, in the UK, non-(EU+UK) students are 27% in CB vs 34% in biology and 42% overall. See Fazey et al. (2005a,b); Wilson et al. (2016); Melles et al. (2019) for a detailed discussion of issues of workforce and research-output diversity in conservation biology.

animal species per year after program launch. These increases are of meaningful scale relative to baselines of 14 countries, 58 plant species, and 151 animal species. Critically, this expansion is geographically targeted and does not come at the cost of quality. Papers on linguistically dissimilar countries (LD papers) rise by around 19 per year. The effects are twice as large for publications in multidisciplinary and Q1 biology journals than in lower-ranked venues. Research on linguistically similar countries (LS papers) shows no corresponding increase, ruling out a general uplift in research productivity. Decomposing by authorship reveals that both locally-connected and migrant researchers shift their agendas after ETMP introduction. Migrant researchers disproportionately drive new work on species and habitats exclusive to their home regions, while locally-connected researchers expand toward universally distributed species. Importantly, new research does not crowd out existing work on nearby geographies, suggesting that incoming migrants expand the frontier of scientific attention rather than redirect it.

These results suggest that researchers coming from new geographies possess distinct skills and preferences from the incumbents. To explore this, I link over 300,000 articles to their research sites (the habitat of the studied organism) and compare papers written at the same university, published in journals of similar rank, in the same year. Three findings emerge. First, papers by migrant researchers are disproportionately sited in physically harder-to-reach locations, measured by road density around the field site. This directly alleviates roadside bias, the well-documented tendency of biological research to concentrate on organisms found in accessible areas near roads and settlements and one of the most persistent sampling distortions in conservation science (Kadmon et al. (2004); Hughes et al. (2021)). Second, their research focuses on countries where the policy need for scientific evidence is most acute. These are sites located in countries with lower environmental performance scores and weaker democratic institutions, both of which are important predictors of an inadequate scientific evidence base in the climate and conservation literature (Amano and Sutherland (2013)). Third, and most strikingly, researchers who bear these additional costs receive no citation reward for doing so. If anything, they face a slight citation penalty relative to peers publishing in the same outlets in the same year. Since the comparison controls for publication venue and year, this penalty cannot be attributed to differences in research quality. Taken together, these results imply that the geographic reallocation of science documented in this paper is driven by supply-side preferences that operate against the grain of demand-side incentives.

1.1 Related literature

This paper makes several contributions to the literature, most importantly the existing work on geographical orientation of scientific enterprise. The production of scientific knowledge is highly concentrated in a small number of universities located in developed countries (Nagaraj and Yao (2026)), whose researchers play a disproportionate role in determining the questions being asked and populations being studied. Consequently, de-

veloped countries and their priorities also receive disproportionate focus in a variety of science and social sciences, as discussed in Section 2. A growing literature suggests that scientific attention is strongly shaped by the accessibility of research sites. In agricultural science, for example, even modest travel costs can influence the choice of studied crops (Kantor and Whalley (2019); Moscona and Sastry (2025); Akerman et al. (2025); Andrews and Smith (2025)). Here, geographical proximity (and its correlates, like informational advantages) are so influential that research effort becomes concentrated on crops grown near a university, despite other crops being more economically important to the broader region. This paper, instead shows that the composition of research workforce is an important determinant of the geography of scientific attention as educational migration can partially offset for proximity premium which shape research agendas.

Thus, I also contribute to the burgeoning literature on the centrality of migrants in scientific workforce in developed countries (Hunter et al. (2009); Moser et al. (2014); Ganguli (2015); Becker et al. (2024)). Alongside papers from inventors outside science (Ager and Brückner (2013); Akcigit et al. (2017); Arkolakis et al. (2020); Posch et al. (2026)), the key focus of the literature on migrant researchers is on the importance of diversity in doing innovative and highly cited research, as opposed to bridging research gaps on questions of concern to origin countries. In this context, the closest papers to ours are Subramani and Fry (2025) and Ash et al. (2022). In Subramani and Fry (2025), the researchers isolate a mechanism through which field and topic choices made by early-career researchers directly reshape the geography and content of scientific research. But here, migrants receiving postdoctoral positions in another (developing) country are more likely to work on health topics associated with the country they move to, re-emphasizing the distance penalty from literature on agriculture. On the other hand, Ash et al. (2022) focuses on return migrants to China but does not study effects on direction of science defined by geographical metrics.

Outside the literature on migrants, a few papers look at homophily between research background or social milieu and the direction of research, for example, on gender (Koning et al. (2021)), race (Dossi (2024)) and socio-economic circumstances (Einiö et al. (2019); Truffa and Wong (2025); Keding and Ritterrath (2025)). I extend this literature by focusing on geography, and importantly highlighting the channels driving homophily could encompass comparative advantage and preferences.

The paper also contributes to the literature on the internationalization of higher education in Europe and Asia through the expansion of English-taught master's programs. While a large body of economic research documents the role of language and medium of instruction in shaping trade flows, migration, and capital movements (Melitz (2008); Ku and Zussman (2010)), comparatively little is known about the consequences of linguistic convergence and anglicization within developed-countries. The closest paper to mine studies the introduction of English-taught programs in Italy (Nocito (2021)), al-

though some do examine the expansion of Erasmus programs across countries (Parey and Waldinger (2011); Parey et al. (2017)). These papers primarily focus on labour-market outcomes of studying in ETPs or studying abroad. More indirectly, Avdeev (2025) studies how cross-border educational migration between Netherlands and neighboring countries affects long-run outcomes such as international social ties and residential mobility. This paper extends that literature by introducing a new outcome—namely, the direction and geographic relevance of scientific research—and by providing causal evidence on how educational internationalization shapes scientific production. In doing so, it contributes empirical grounding to discussions of scientific diplomacy and international knowledge production that have largely remained rhetorical and outside the core economics literature.

The rest of the paper proceeds as follows. Section 2 describes the institutional background underlying the expansion of English-taught master's programs. Section 3 describes the data and measurement strategy. Section 4 outlines the research design and identifying variation. Section 5 presents the main empirical findings and discusses their implications.

2 Background

Across disjointed subsections, I present key background information necessary to conceptualize the results of this paper. The scope of the problem – geographical skewness of science at large and biological sciences specifically – are discussed in Section 2.1, followed by a primer on the expansion of ETPs in non-anglophone countries (Section 2.2). The final subsection discusses of organization of scientific and mentoring relationships within conservation biology, which shall be helpful in discussing external validity of the results.

2.1 Geography of Science

The geography of production and relevance of science are extremely skewed as developed countries are over-represented on both fronts, compared to their shares of world population and economic output. Recent statistics from Nagaraj and Yao (2026) suggest that one-in-four papers are written by US-based authors and one-in-three papers are written in the EU. Their contribution to top journals and publications is even starker - 80% of top permille papers are produced by scientists in the US and EU. Several structural advantages possibly contribute to their dominance in scientific production. Presence of exceptional universities, ability and openness to attract global talent, and generous scientific funding and endowment create an environment for production of research.

The geography of scientific relevance is also heavily tilted towards topics of interest to developed countries. Global, cross-science estimates are not available, but I document the

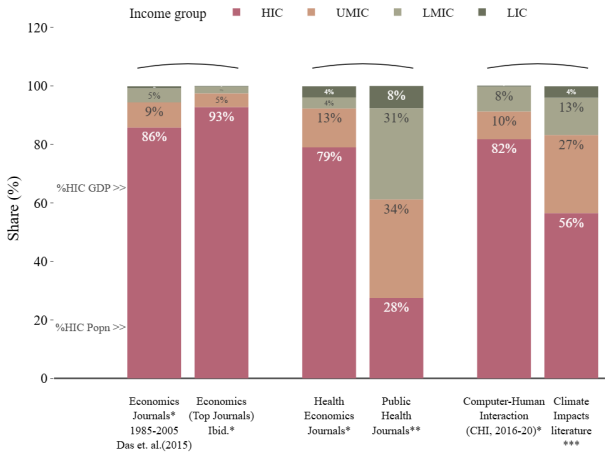


Figure 1: Geography of Relevance of Science Across Fields

Notes: This figure shows the share of papers relevant to countries at different levels of economic development. Data for economics-related publications comes from Das et al. (2013), which use mentions of names of countries in title and abstract to quantify research relevance and spans 1985-2015. For health economics, I use Pitt et al. (2016), which covers 2010-2014. Both classify countries based on incomes at the time of publication. The remaining bars use income-classification from 2023. In case of public health journals, data on disease-level paper counts comes from Schmallenbach et al. (2025), which is then linked to mortality data from GBD 2023 to arrive at the distribution of relevance for papers since 1999. For human-computer interaction, data comes from publications between 2016-2020, linked to country of relevance by Linxen et al. (2021). Finally, the crosswalk from publications to relevant countries for climate-change impacts comes from Callaghan et al. (2021).

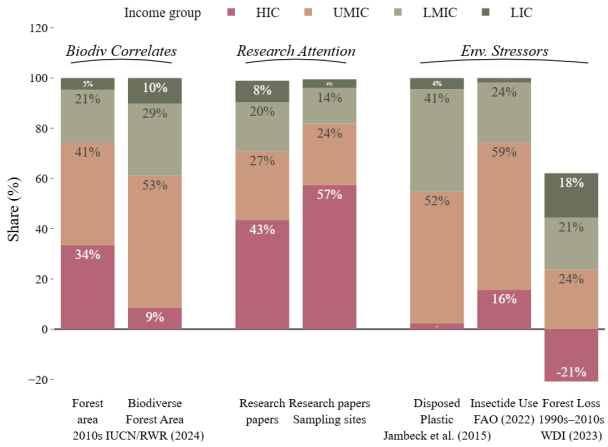


Figure 2: Research effort and need mismatch in Conservation Biology

Notes: This figure plots the share of scientific research relevant to countries along the development spectrum, and benchmarks this against the share of biodiversity found in these areas and the shares of threats to biodiversity concentrated therein. Forest area comes from World Development Indicators (WDI), published by the World Bank. Share of highly-bio-diverse forest is derived using an index from IUCN, published in Albuquerque and Beier (2016). Shares of research papers are calculated using data from Openalex.org and the methods discussed in Section 3. Data on environmental stressors comes respectively from Jambeck et al. (2015), Shattuck et al. (2023)/FAOSTAT and WDI.

relevance of science in a variety of fields of different vintages in Figure 1. The methodology used to compute the shares differs by field under consideration and is discussed in the note beneath the figure.

The key takeaway is the dominance of HICs in research relevance, with disproportionately large attention in both older fields like economics and younger fields like impacts of climate change and Human-Computer Interaction, suggesting that historical advantage alone is insufficient in explaining disparate outcomes. Their dominance in public health seems lower, but it is driven down by evolving global disease-burden transitions, where many LICs and LMICs now face a growing burden of non-communicable diseases. Thus, the large body of work conducted on NCDs in developed countries becomes "relevant" for LICs, even though contextual factors like poorer health systems and greater reliance on non-allopathic medicine might limit their applicability.

A comprehensive review of factors driving research skewness is beyond the scope of this paper. It is important to emphasize, however, that returns to research on social and scientific questions might be higher in developing countries, but stakeholders therein often do not have resources to mobilize scientists to study their problems, nor do they have the institutional abilities to signal their priorities. The endeavor to identify geographical gaps in research understates the scale of these gaps if developing countries do not just require answers to universal questions in their context, but also have context-specific questions which might not arise in developed-country settings.

2.1.1 Geography of Biological Knowledge

The information on the geography of relevance of biological knowledge is condensed into Figure 2. Biodiversity itself is disproportionately located outside high-income countries: only 9% of biodiverse forest area (IUCN/RWR) lies in HICs, compared to 53% in upper-middle-income and 29% in lower-middle-income countries. Yet research attention is heavily skewed toward richer countries: 57% of research sampling sites and 43% of research papers focus on HICs, while low-income countries account for just 4–8%. Environmental stressors, by contrast, are concentrated in middle-income countries—for example, 59% of insecticide use and 52% of mismanaged plastic waste occur in UMICs, and forest loss is also disproportionately borne by non-HIC regions. Taken together, the figure highlights a systematic mismatch as biodiversity and environmental pressures are concentrated in area distinct from countries where scientific attention remains centered.

2.2 English-Taught Master's Programs

The last few decades have witnessed a sharp increase in educational migration to non-traditional destination countries—that is, countries outside the United States, the United

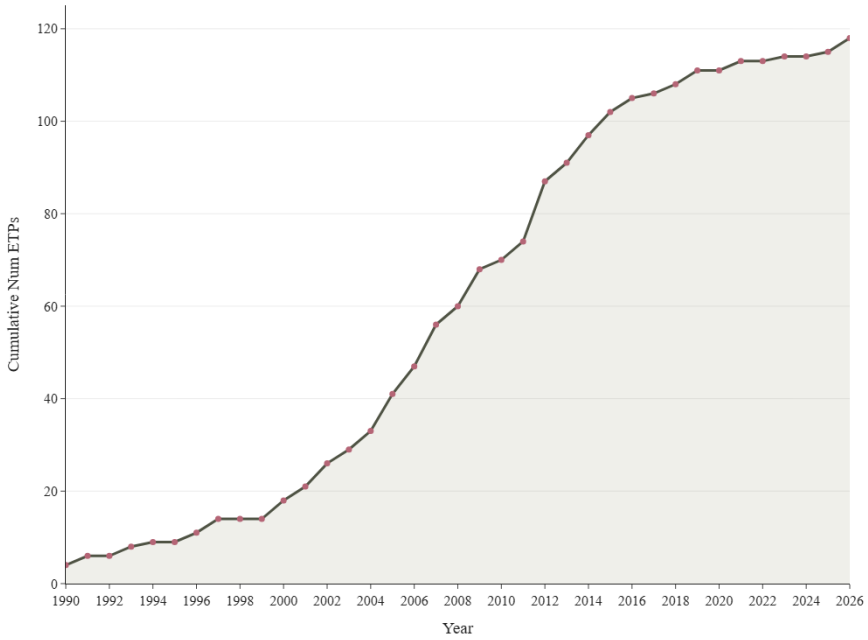


Figure 3: Growth of biology-relevant ETMPs at sample universities (1990-2026)



Figure 4: Non-Anglophone countries with sample universities

Notes: Countries in Sample (N=20) are in shades of cream and blue, with color intensity reflecting the number of universities included in the sample (N=118). The progress circles show share of ETMPs launched in different decades for universities in that country.

Kingdom, Canada, and Australia.[4] A growing share of these flows has been absorbed by non-Anglophone, high- and upper-middle-income countries in Europe and Asia, which constitute the focus of this paper (Figure 4). By 2019, the 20 countries in my sample attracted over 2.5 million student migrants, comparable in magnitude to the four premier international education destinations (2.9 million students). A central mechanism through which these countries expanded access to international students is the introduction of English-taught master's programs (ETMPs): one- to three-year professional degrees offered in fields with broad international demand. By adopting English as the medium of instruction, these programs lowered language barriers that had previously restricted access to elite universities in non-Anglophone countries, thereby broadening the pool of prospective international students (see Figure 3).

Several institutional forces shaped the timing and location of ETMP adoption. First, the Bologna Process, initiated in 1999, sought to harmonize higher-education systems across Europe by standardizing degree structures and introducing the European Credit Transfer and Accumulation System (ECTS). Prior to Bologna, limited compatibility of degrees across countries constrained both educational and labor mobility, as credits were difficult to transfer and employers faced challenges in interpreting foreign credentials. The Bologna reforms increased cross-border comparability and encouraged universities to redesign curricula. Many institutions used this transition as an opportunity to introduce English-taught programs aimed at attracting a broader international student base. Together with the Erasmus programs, Bologna represents a major phase of higher-education internationalization among universities in the sample.

Financial pressures also played an important role. Following the 2008 global financial crisis and the subsequent eurozone crisis, constraints on public funding increased universities' reliance on own-source revenues, particularly international student fees. English-taught professional master's programs provided a scalable means of attracting fee-paying international students, contributing to a shift away from exclusive dependence on government transfers.

Finally, two longer-run trends reinforced internationalization pressures. First, demographic shifts in several countries—aging populations and declining domestic student cohorts—reduced the supply of national students and early-career researchers, particularly in countries such as Japan. Second, global university rankings increasingly incorporated indicators of internationalization, often measured as the share of international students enrolled. In response, governments introduced funding schemes that rewarded universities for meeting enrollment targets for foreign students. These policies were largely discipline-neutral, leading to rapid expansion of English-taught programs in fields with large international markets, such as computer science and business administration. Biology and related disciplines, while present, remained a secondary focus of these interna-

[4]The international education industry refers to these countries as the "big four" destinations.

tionalization efforts.

2.3 Organization of research in conservation biology

Modern conservation science academia has four features which make it an ideal setting for my research. Some of these features permeate science at large, and while the precise timing of the results hinges on these features, underlying theoretical dynamics might be more diffuse across science and social sciences.

First, fieldwork and primary data collection is central to the discipline. While some researchers in conservation biology now rely on large-scale secondary datasets generated through citizen-science efforts, government surveys and camera-traps, much of the frontier research focuses on collecting and using samples from diverse habitats like tropical rain-forests and deserts. Thus, graduate students need to be willing and able to brave field sites both near home and away, and have inclination to learn skills (like language, cultural sensitivity) which may not be directly related to the science involved in their work. Prior experience with these field sites is, thus, prized. It is fairly common for undergraduate students interested in academic pursuit to serve as interns and survey staffers for larger teams foraging through forests for their samples.

Second, researchers within the field increasing frame their agendas as being focused on broad categories of organisms, methodological innovations, or aspects of life-cycle like reproduction or excretion. This leaves sufficient room for incoming graduate students to leverage existing knowledge of the lab, but apply it to new geographies. Tables 3 and 4 provides an overview of all conservation biology labs at Cambridge and Princeton University, specifically highlighting the vast geographical scope of the labs. Most labs are either centered around policy questions or cater to species found in dozens of countries.

Third, publication timelines and expectations in conservation science are relatively streamlined. Papers are short, report few results and remain under review for less than half a year. Masters theses are usually published within a year from being written. Thus, effects of diverse researchers on outcomes could arise pretty quickly.

Fourth, attention issues could be seriously impeding research within conservation. There are over a million named plants and animals, and it is not likely for individual researchers to know many of them. Diverse individuals could bring in knowledge about organisms from their niches.

3 Data: Sources & Descriptive Statistics

This project relies on four key streams of data, which are discussed in turn.

3.1 Outcome Variables

3.1.1 Relevance of research output

I collect the metadata, including the title and abstracts, for over 0.8 million papers from 1300 journals of biological and agricultural sciences and general-interest journals like *Nature* and *Science*. The data is scraped using an API from Openalex.org, a website containing information on millions of published papers and widely used in meta-science research (for example, Arts et al. (2025)). To restrict analysis to the realms of conservation biology, especially for papers published in general-interest journals, I remove the publications whose topic - as classified by Openalex - differs markedly from fields under consideration.[5]

The titles and abstracts contain valuable information to ascertain the geographical relevance of the papers. Notably, I parse these textual sources for three indicators - names of countries, names of GAUL-1 units (sub-national administrative units or states), and plants and animal names. I also assemble a unique and globally representative data on habitats of over a million species to link each organism to countries where it is found (discussed in Section 3.1.2). Using these indicators, I can assign each paper to all countries it could plausibly be relevant for. A paper on insects, for example, could be relevant to each of the 180 countries whereas that on tigers could be plausibly relevant for less than 10 countries.
This is aggregated at the university level. Thus, I know the number of papers of relevance to country written at the sample university in a year. High endemism and slow-changing habitats of animals help provide precise measures of geographical relevance[6]

3.1.2 Habitat ranges of studied species

To map the biological focus of papers to geographical areas, I assemble data on the habitat ranges of plant and animal species from multiple biodiversity databases. The primary source is the Global Biodiversity Information Facility (GBIF), which aggregates biodiversity records from citizen scientists, taxonomists, governments, museums, and intergovernmental organizations such as the IUCN. Focusing on the plant and animal kingdoms, I harmonize species-location information across these sources using ISO-3 country codes, including marine species observed within countries' Exclusive Economic Zones (EEZs).

[5]Despite its depth and comprehensiveness, Openalex data suffers from a key limitation. Around half the articles do not have an associated abstract in the dataset. A large share of these articles are from Elsevier, who restrict the use of their abstracts by open-source aggregators. I use LSE's subscription to Scopus to backfill missing abstracts using standard procedures. The sample used here - 0.8 million papers - are those with complete metadata.

[6]It might be harder to construct a similar index for papers in other fields, like economics. For example, even when a paper on, say minimum wages, might be based on data and policy from one country, it is uncertain whether countries with or those looking to instantiate minimum wages benefit from the paper. Similarly, in public health, demographic and epidemiological transitions might create a wedge between relevance in the present and lasting relevance of research.

GBIF provides broad global coverage, but coverage quality differs across taxa and regions. In particular, information on plant habitats is less comprehensive than for animals. To improve coverage of plant distributions, I supplement GBIF with the World Checklist of Vascular Plants (WCVP), a widely used and regularly updated database maintained by botanical researchers (Govaerts et al., 2021).

Country-level species ranges are useful for identifying the broad geographical relevance of research, but they are too coarse for measuring ecological and logistical characteristics of habitats, such as climate, remoteness, or terrain. I therefore construct finer-grained habitat measures using spatial occurrence and range data. First, I use GBIF's standardized occurrence product, which records species observations on a global hexagonal grid of roughly 36 km^2. Because occurrence records may reflect known biases in citizen-science and observational data collection (Dickinson et al., 2010; Geldmann et al., 2016; Roberts et al., 2016), I supplement these data, where available, with expert-curated range maps from the International Union for Conservation of Nature (IUCN) and *Birds of the World*.[7] I subsequently merge these geographical datasets with global gridded data on climate, economic activity, land use, road networks, and population density to construct measures of fieldwork difficulty and habitat accessibility.

3.1.3 Researcher composition

A key intermediate outcome in this project is the effect of ETMP on composition of young researchers at these universities. While the names and affiliations of all first and last authors of papers are available, their nationalities are not known. Using a machine-learning-based prediction algorithm, I assign the name's probability of belonging to persons with select nationalities. The algorithm is far from perfect, but provides reasonable predictions useful for a sniff test. For example, it assigns most Anglo-Saxon names to United Kingdom, and does not assign a single author to Canada, Australia or the United States. Importantly, though, it is a good predictor of the linguistic origin of the author even though its ability to match nationalities might be severely limited. Thus, for university-year, I compute the number of authors coming from linguistically similar and dissimilar countries as the primary first-stage outcome variables.

3.2 Identifying Variation

The variation in my project comes from the staggered opening of English-taught programs across universities in my sample. There is no consolidated dataset of launch dates of ETMPs. So, I artisanally collect and verify this data through multiple sources. Most importantly, I go through the internet-archive websites of each of the universities over several years and identify the announcements associated with opening of different programs at the university. Among the newly opened programs, I single out those run at

[7] See: https://birdsoftheworld.org/bow/home

the biology or agriculture departments, taught in English, and open to international students without constraints. For the sake of treatment definition, I retain the first of these programs launched at the university.

3.3 Covariates

The paper relies on a series of covariates from geographical and socio-economic data-sources. Economic variables like country-level per-capita GDP and Gini indices are sourced from World Bank's World Development Indicators (WDI), alongside remotely sensed and globally harmonized climate variables from standard satellite sources like Fick and Hijmans (2017) and Block et al. (2024).

3.4 Summary Statistics for Outcome Variables

I provide an overview of summary statistics for key outcome variables used in the analysis, separately for control universities in Anglophone countries and summaries from before and after the launch of ETMP for treated universities. Table 1 focuses on researchers whereas Table 2 is about research papers and features of corresponding research sites. A key concern with these summary statistics, as with all summaries of event-study designs over long periods, is that they conflate variation from treatment with changes over time. Thus, I limit interpreting any of these statistics very closely. Directionally, however, they provide support for the underlying hypotheses, which I discuss in greater detail with appropriate methodology later.

I note three things about Table 2. First, this table provides summary statistics about the research site of over 0.3 million papers in my sample across multiple dimensions. Second, across indicators, standard deviations are very large, partly because of the noise in inferring sampling sites from text of title and abstract and also because of diversity in research sites for conservation biology. In case of precipitation, for example, both desert areas and high-rainfall rainforests are potential targets of research. Third, universities in non-anglophone countries have stronger research ties to sites which are in accessible areas than those in anglophone countries.

4 Research Design

To identify the effects of changes in the diversity of the research workforce on the direction of scientific research at the university level, I exploit a plausibly exogenous shock to the composition of early-career researchers induced by the staggered introduction of English-taught master's programs (ETMPs) at universities in non-Anglophone countries. The rollout of ETMPs lowered language barriers to entry for prospective international

Table 1: Researcher Diversity Outcomes by Career Stage

Indicator	Control	Treated (Pre)	Treated (Post)
Early Career: Share LD	0.272	0.245	0.359
	(0.151)	(0.171)	(0.188)
Early Career: Distinct Nationalities	15.675	11.346	20.644
	(11.200)	(7.718)	(11.004)
Mid Career: Share LD	0.289	0.335	0.456
	(0.182)	(0.236)	(0.252)
Mid Career: Distinct Nationalities	11.148	7.387	13.234
	(9.251)	(5.259)	(8.131)
Late Career: Share LD	0.171	0.385	0.433
	(0.145)	(0.242)	(0.259)
Late Career: Distinct Nationalities	8.524	6.464	10.484
	(6.864)	(4.103)	(5.658)

Notes: Entries report means. Standard deviations are reported in parentheses below each mean. The ratio outcome is the share of newly entering researchers whose nationality is linguistically different from the university's primary language. Distinct nationalities counts the number of unique ISO3 nationalities among newly entering researchers in a university-year.

students and researchers, generating variation over time and across institutions in the linguistic and geographic origins of research talent. Treated universities are matched to a panel of never-treated universities located in Anglophone countries, which historically attracted large shares of international students but did not experience comparable changes in language of instruction during the study period.

I estimate dynamic treatment effects using an event-study framework following Callaway and Sant'Anna (2021) that compares outcomes at treated universities before and after ETMP introduction to contemporaneous outcomes at matched control universities. The specification used is as follows:

$$Y_{iut} = \alpha_u + \alpha_t + \sum_{g \in \mathcal{G}} \sum_{j=-5}^{10} \beta_{j,g} \, 1\{t - T_u = j\} \mathbf{1}\{u \in g\} + \varepsilon_{iut}.$$

Reported estimated coefficients from Sun & Abraham (2021) using csdid package:

$$\beta_j = \sum_{g \in \mathcal{G}_j} w_{g,j} \, \beta_{g,j}$$

The primary outcomes include measures of researcher composition, which is proxied by the linguistic similarity between first authors' predicted countries of origin and the host university's primary language, as well as measures of the geographic relevance of research output. Estimation follows modern difference-in-differences methods that accommodate staggered treatment timing and heterogeneous treatment effects across cohorts.

Table 2: Characteristics of Research Paper's Field Site

Indicator	Control	Treated (Pre)	Treated (Post)
Highways (m/km^2)	9.02	10.49	10.92
	(45.21)	(55.83)	(53.28)
Primary Roads (m/km^2)	35.55	67.92	59.99
	(117.04)	(157.74)	(143.58)
Secondary Roads(m/km^2)	77.87	95.61	90.78
	(165.15)	(145.14)	(143.16)
Tertiary Roads (m/km^2)	99.31	161.69	150.86
	(357.05)	(315.36)	(319.49)
Local Roads (m/km^2)	148.11	186.64	175.82
	(776.14)	(738.95)	(666.25)
Population Density (km^2)	88.05	186.68	166.29
	(907.33)	(1145.89)	(965.31)
EPI Score (max 100)	67.20	68.99	67.77
	(12.36)	(12.84)	(13.53)
Deliberative Democracy (max 1)	0.658	0.637	0.612
	(0.232)	(0.254)	(0.272)
Egalitarian Democracy (max 1)	0.607	0.604	0.579
	(0.225)	(0.256)	(0.273)
Liberal Democracy (max 1)	0.663	0.630	0.602
	(0.251)	(0.275)	(0.297)
Participatory Democracy (max 1)	0.523	0.502	0.480
	(0.183)	(0.209)	(0.225)
Electoral Democracy (max 1)	0.742	0.710	0.680
	(0.233)	(0.265)	(0.289)
Annual Precipitation (mm)	291.92	525.51	617.12
	(2046.88)	(1661.13)	(1502.80)
N	121,924	66,355	124,205

*Notes:*Entries report means. Standard deviations are reported in parentheses below each mean. The variables provide information on the characteristics of the research sites inferred from the title and abstract of the research paper. A variety of data sources are used to gather globally gridded datasets corresponding to each measure.

Identification relies on a parallel trends assumption: in the absence of ETMP introduction, outcomes at treated and control universities would have evolved similarly over time. For specifications based on linguistic similarity, this assumption implies that research conducted by non-Anglophone universities on linguistically dissimilar (Anglophone countries) countries would have followed trends comparable to research conducted by Anglophone universities on the territories dissimilar to them (non-Anglophone countries), absent the policy change. Thus, we are assuming that research across different contexts evolve parallely in the absence of the treatment.

This design has both advantages and limitations. On the one hand, comparing trends across distinct geographic contexts helps mitigate concerns about spillovers or strategic interactions among universities operating in closely related research fields or national systems. On the other hand, the design may be sensitive to global "attention" shocks—periods in which certain regions or topics receive increased scientific attention across institutions for reasons unrelated to ETMP adoption. Such shocks could generate differential trends in geographic research focus that are correlated with treatment timing but not caused by changes in researcher composition.

I address these concerns in several ways. First, the event-study specification allows for visual and statistical inspection of pre-treatment trends, and robustness checks following Rambachan and Roth (2023) explicitly assess sensitivity to deviations from parallel trends. Second, the inclusion of university and year fixed effects absorbs time-invariant institutional characteristics and common shocks affecting all universities. Third, in subsequent analysis, I could extend the design to a triple-difference framework that exploits within-university variation across linguistic similarity and research geography, further isolating the role of researcher composition from broader shifts in global scientific attention.

Together, this research design leverages institutional policy variation in graduate education to credibly identify how changes in the composition of early-career researchers shape the direction and content of scientific research.

4.1 Research-site characteristics

To examine whether the composition of researchers changes the types of field sites studied, I estimate paper-level event-study regressions linking each article to characteristics of the research site associated with the species or location studied in the paper. The main outcome in this specification is an index of physical accessibility, constructed as the first principal component of road density around the research site. Specifically, I use measures of highway, primary, secondary, tertiary, and local road density around each site and extract the first principal component. Higher values of this index correspond to more road-dense, and therefore more physically accessible, research sites.

$$Y_{pujt} = \alpha_{uc} + \alpha_j + \alpha_t + \Gamma' X_p + \sum_{k \in \mathcal{K}} \beta_k \left(Treat_u \times \mathbf{1}\{t - T_u \in k\} \right) + \sum_{k \in \mathcal{K}} \delta_k \left(Treat_u \times \mathbf{1}\{t - T_u \in k\} \times Diff_p \right) + \varepsilon_{pujt}. \tag{1}$$

Here, p indexes papers, u indexes universities, j indexes journals, and t indexes publication years. The outcome Y_{pujt} denotes a characteristic of the research site associated with paper p. In the baseline specification, this is the first principal component of road-density measures around the site. $Treat_u$ is an indicator equal to one for universities that ever introduce an English-Taught Master's Program (ETMP), while T_u denotes the year of first ETMP introduction. Event time is grouped into two-year bins, with $\mathcal{K} = \{-10, -8, -6, -4, 0, 2, 4, 6, 8, 10\}$ corresponding to the intervals $[-10, -9]$, $[-8, -7]$, $[-6, -5]$, $[-4, -3]$, $[0]$, $[1, 2]$, $[3, 4]$, $[5, 6]$, $[7, 8]$, and $[9, 10]$. The omitted category is the two years immediately preceding ETMP introduction (-2 and -1).

$Diff_p$ is an indicator equal to one if the paper is authored by a researcher from a linguistically dissimilar country. The specification includes university-by-category fixed effects (α_{uc}), journal fixed effects (α_j), and publication-year fixed effects (α_t). The vector X_p contains controls for the biological kingdom of the studied organism and the type of location associated with the observation. Standard errors are clustered at the university level.

The coefficients β_k capture changes in site accessibility following ETMP introduction for papers not authored by linguistically dissimilar researchers. The coefficients of primary interest are δ_k, which measure whether papers authored by linguistically dissimilar researchers shift toward different research sites following ETMP introduction. A negative value of δ_k implies that these researchers disproportionately conduct research in less road-dense and therefore less physically accessible locations relative to other researchers operating in the same institutional environment.

5 Results

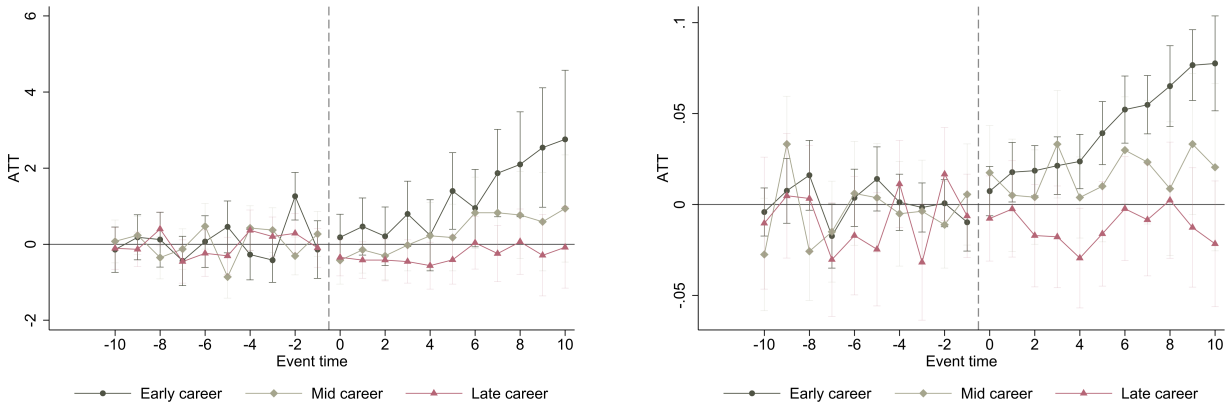
The following five subsections present key results from this project thus far. I start with the first-stage results - estimated effects of the launch of ETPs on diversity of research workforce at the universities - in Section 5.1, followed by effects on research direction in Section 5.2.

5.1 Researcher Diversity

The key first stage results are presented in Figure 5 and corresponding Tables 5 & 6. I find a large, positive and statistically significant increase in two dimensions of researcher diversity – the number of distinct nationalities of researchers and (speaking directly to the mechanism) the share of new researchers coming from linguistically dissimilar countries. In both cases, results are concentrated among early career researchers (ECRs), that

is, those within five years of having written their first paper. Pre-treatment coefficients are small and statistically indistinguishable from zero. I estimate that ECRs come from 1.23 additional countries in the decade following ETMP's introduction, with effects peaking at 3 countries in the tenth post-ETMP year. This is a an 8.2% increase over the baseline for ECRs - roughly 15 different countries of origin.

The share of incoming ECRs from linguistically dissimilar countries also shows positive and statistically significant increase after the launch of ETMP. From a baseline of 27%, there is 5 percentage point estimated effect or roughly a quarter higher in post-period. As with the last outcome, the results for mid and late-career researchers are noisier, not consistently positive, and not statistically significant. The lack of a clear trend in diversification of senior researchers helps rule out department-wide internationalization packages which might combine the launch of new master's programs with diversification of hiring practices, additional funds, or the creation of international research portfolio. While omitted here, the effects on number of hires at mid- and senior- career levels (irrespective of nationality) are not statistically or economically significant.



(a) Number of countries **(b)** Share of LD authors

Notes: Figure (a) shows event study estimates for the number of different nationalities for incoming researchers, split across career stages. ATT for ECRs=1.23 countries/year (p-val=0.002). Figure (b) shows the share of linguistically dissimilar (LD) researchers in the incoming cohorts, split across career stages. ATT for ECRs = 5 pp (p-val=0.001)

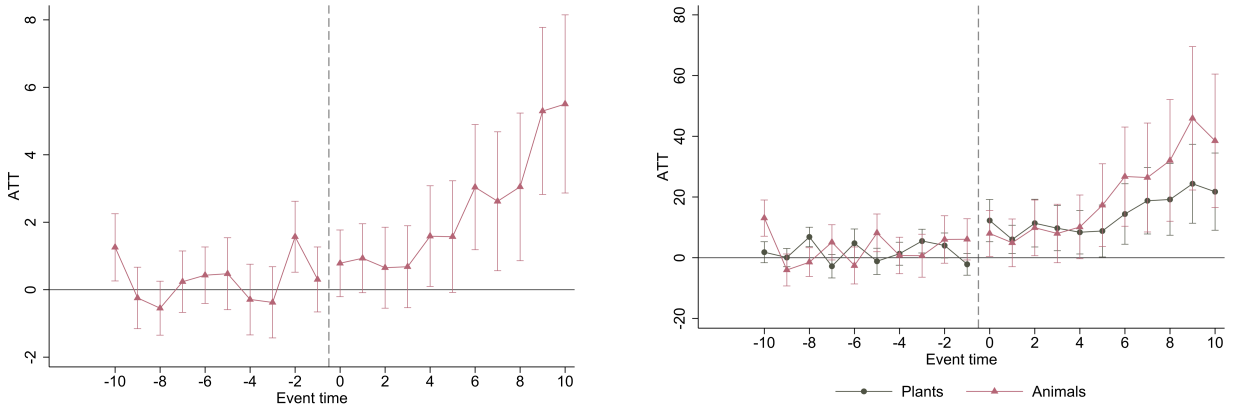
Figure 5: Estimates for Effect of ETMP on Geography of Researchers

5.2 Research Diversity

Figures 6 and 7 present the main results on changes in the geographic relevance of research after the launch of ETMP. Four outcomes are explored. On the extensive margin, Figure 6 plots the coefficients for estimated effects of ETMP on number of countries and

organisms studied at the university. The estimates are positive, statistically significant, rising over time, and economically important. Compared to the baseline, researchers in anglicizing universities study 2.3 additional countries and over 14 different plants and 20 different animals in a given year. The estimates are larger for animals than plants as the fieldwork difficulty for animal-related research is often more stringent than plants. Baseline averages for these outcomes are 14 countries, 58 species of plants and 151 animals, respectively.

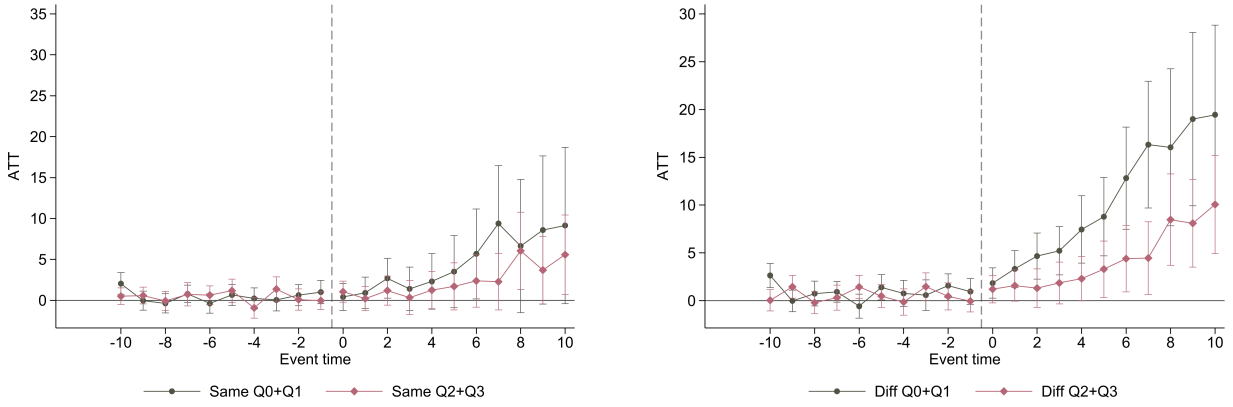
Figure 7 provides evidence consistent with the mechanism that new research is concentrated among habitats in linguistically dissimilar countries. Panel (a) plots the estimated effects of ETMP on research in linguistically similar countries, split by the ranking of journal of publication. The reported effects are small (around 4 additional papers a year for high-quality journals) and not statistically significant at conventional confidence levels. However, for linguistically dissimilar countries, the ATT over the following decade is around 19 additional papers a year, which is statistically and economically significant. The estimates for high-quality journals – multidisciplinary science journals (e.g. *Nature*, *PNAS*) and Q1 biology journals – are twice as large as lower-ranked journals. The coefficients also increase for subsequent years, in line with accumulation of talent and lags in publications and development of research agendas. Throughout, the pre-treatment coefficients are small and statistically indistinguishable from null effects. However, when pooled into 3-year buckets, as in Table 7, they do appear statistically significant. Despite that, the magnitude is miniscule (at around 1 paper) and there is no clear upward trend as the coefficient for earlier and later period are almost identical. In later drafts, I shall ameliorate this issue with a synthetic difference-in-differences analysis and better selection of control group.



(a) Number of Countries Studied **(b)** Number of Organisms Studied

Notes: This figure presents event-study estimates for geographical (a) and taxonomic (b) expansion of research frontier at treated universities, compared to the controls. *Author's calculations*

Figure 6: Research Diversification - Extensive Margin Outcomes



(a) Papers on Linguistically Similar Countries

(b) Papers on Linguistically Dissimilar Countries

Notes: This figure shows event study estimates separately for count of papers with geographical relevance to linguistically similar (a) and dissimilar (b) countries. It further separates them by the ranking of the publication outlet
Source: *Author's calculations*

Figure 7: Research Diversification & Geographical Focus

5.2.1 Linking first and second stage outcomes

This subsection decomposes the rise in papers relevant to linguistically dissimilar countries by authorship and degree of relevance. That is, for each university and year, I calculate the number of papers that are exclusive LS-relevant or LD-relevant and those relevant to both set of countries, by each author type. Figure 8 presents the key result.

Three key results emerge. First, the research agendas of both LS and LD authors change after launch of ETMP. Second, while LS authors focus on species which are universally found,[8] LD authors focus on LD-exclusive species and universal species at similar rates, relative to the attention in pre-ETMP period. Finally, there is no statistically significant effect, especially negative effect, on research on LS topics and geographies. Thus, new research does not crowd out knowledge generation about nearby research.

In the bottom panel of the figure, I limit the sample to papers mentioning a country or state where the sampling event took place. Here the results are clearer and sharper. For the Ls authors, changes in sampling composition change, but mildly so. For the LD authors, on the other hand, there is a steep and continuous increase only for sampling in LD countries.

These results point to potentially interesting organizational dynamics that could result from spillovers between LS and LD authors, in line with the effects of diversity on knowl-

[8]This is a slight misnomer. In the context of this paper and measurement approach, a universal specie is the one found both in a linguistically similar and a linguistically dissimilar country from the university's standpoint.

edge and inventive capacity. Importantly, the point estimates for LS authors working on LD paper is half the estimate for Universal papers, indicating some role of geographical familiarity in shaping research choices.

5.3 Robustness - Placebo Test Using Economics Publications

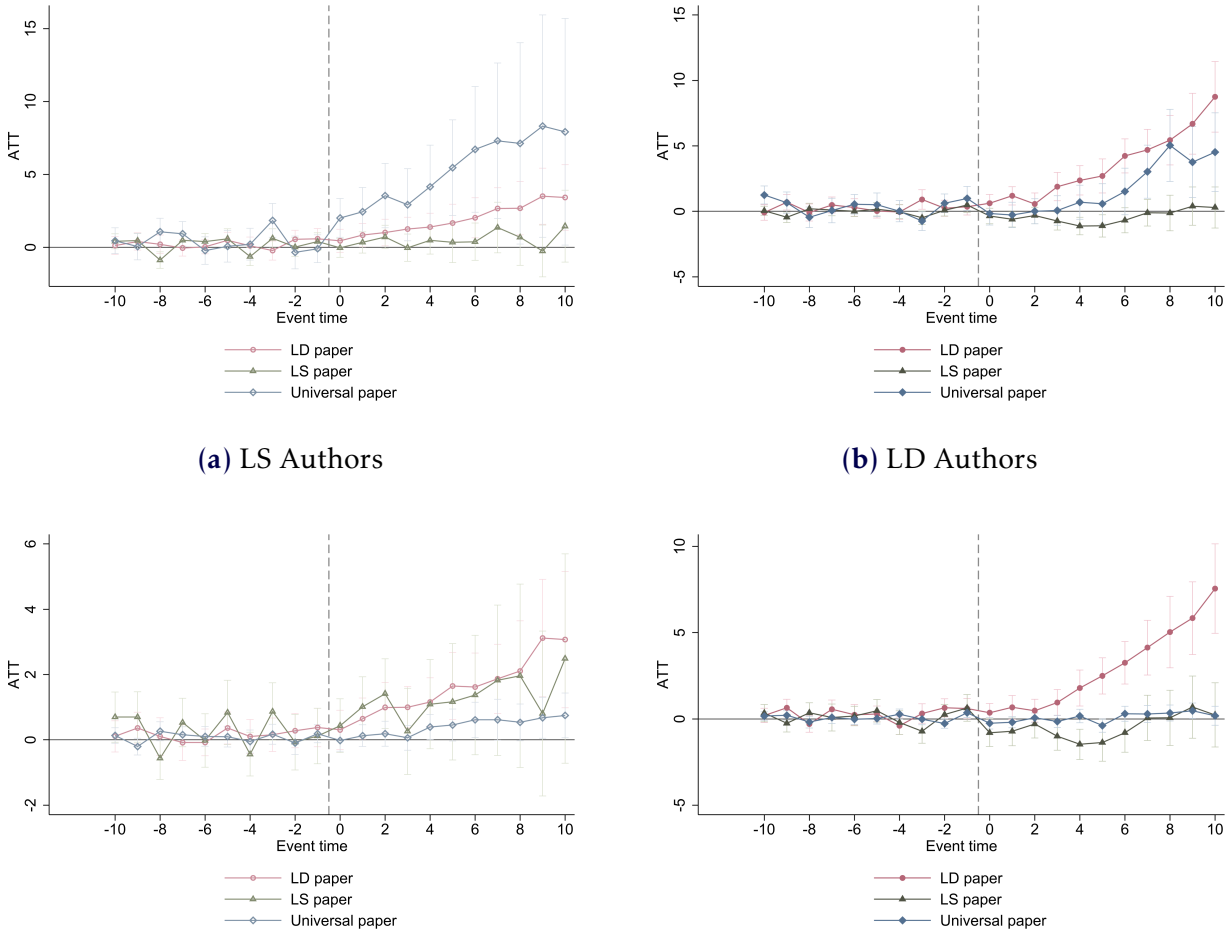
A key concern for identification is that the introduction of ETMPs may coincide with broader university-level internationalization initiatives or fiscal shocks that independently affect research production and collaboration patterns. If so, the estimated effects may reflect these concurrent institutional changes rather than the causal effect of ETMPs themselves. While the absence of meaningful effects on senior hiring and overall hiring volumes in Section 5.1 partially alleviates this concern, such interventions may still affect research activity through other channels.

To address this possibility, I estimate placebo effects on the geography of research in economics publications produced by the same universities. Economics provides a useful placebo setting because spillovers from increased diversity within biological sciences departments to economics research are likely to be limited. Thus, significant effects in economics would instead suggest the presence of broader campus-wide shocks coinciding with ETMP adoption.

Figure 9 reports the results. I find economically small and statistically insignificant estimates throughout most of the post-treatment period, with no evidence of persistent increases in internationally diversified economics research following ETMP introduction. Even in the later post-treatment years, estimated effects remain modest at roughly one additional paper per year, compared to approximately forty papers per year for the main biological sciences outcome. The absence of comparable effects in economics suggests that coincident university-wide shocks are unlikely to account for the magnitude of the main results.

5.4 Mechanisms

So far, I have shown that reducing barriers to participation of LD authors redirects scientific attention towards LD countries. This section provides evidence that LD researchers work in settings that are distinct from their LS counterparts, and are possibly harder to work in. To show this, I collect data on the habitats and occurrences of a wide range of organism featured as objects of research, and create scores of fieldwork difficulty for each article. To do so, I overlay the previously discussed habitats with global gridded datasets of road density, temperature and precipitation, elevation, forest density, closeness to habitable populations, country-level scores of democratic and environmental outcomes, and land use patterns. To show differences in LS and LD authors as suppliers of research, I use Equation 1 and plot estimates of β_k and δ_k in Figure 10.



(a) LS Authors

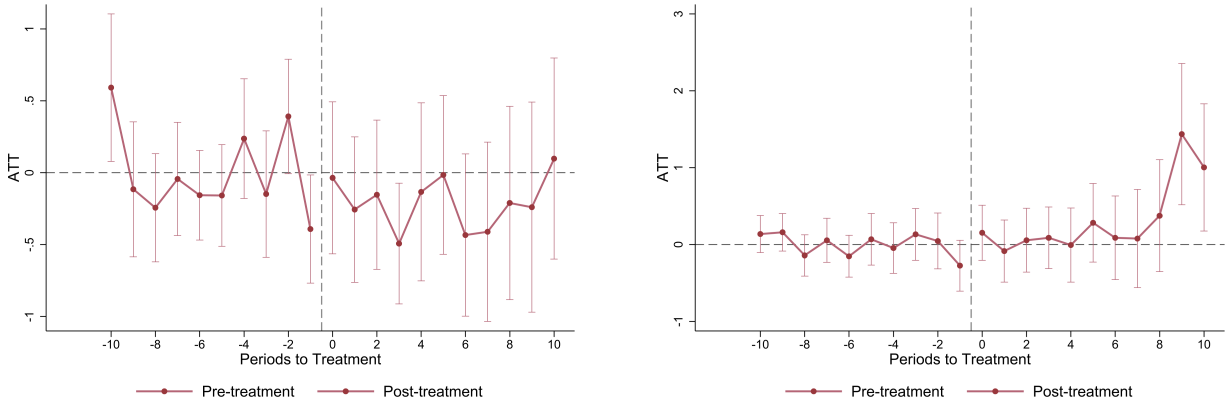
(b) LD Authors

(c) LS-Relevant Papers

(d) LD-Relevant Papers

Notes: This figure decomposes changes in research direction by both author background and geographical relevance of papers. Panels (a) and (b) report event-study estimates separately for linguistically similar (LS) and linguistically dissimilar (LD) authors, further split by whether the paper focuses on LS geographies, LD geographies, or universally distributed species. Panels (c) and (d) report overall effects on LS- and LD-relevant papers by publication quality category. Pre-treatment coefficients are small and statistically indistinguishable from zero throughout.

Figure 8: Research Diversification, Author Composition, and Geographical Focus



(a) Number of countries studied **(b)** Number of LD papers

Notes: Figure (a) shows event study estimates for the number of different countries studied by the economics department after the biology department introduces and ETMP. Figure (b) shows the number of LD papers in economics journal. While it is statistically significant at 1 paper in Year 9, it dwarfs the ATT for biology department, which is 19 papers.

Figure 9: Estimates for Placebo check

Three key results emerge. First, part (a) of Figure 10 shows that LD authors work on in field-sites with lower road density. These areas are less accessible, therefore localized knowledge, informal modes of transport, and potential exposure to extreme landscapes and climates could be larger. Roadside bias – the concept that conservation biologists over-sample and over-focus on research in accessible and reachable areas – is a first-order concern in the field of biodiversity (Kadmon et al. (2004); Hughes et al. (2021)), which these graduate students could bridge.

In addition, parts (b) and (c) of Figure 10 shows that LD authors work on papers creating evidence about less democratic and less environmentally friendly countries. Democratic institution of countries containing the field-work locations is measured using Varieties of Democracy data, and is matched to the year closest to the publication year. For environmental performance, I use the Environmental Performance Index (EPI) developed by researchers at Yale University. These are used here as proxies for policy need as the former is strongly correlated with fewer number of papers about a country whereas the latter could be seen as a direct outcome of incomplete knowledge to regulate local environmental concerns. alongside previous evidence, and in line with observations from field scientists, LD authors possibly bear additional costs in pursuing research in less-conducive settings.

Finally, I document that LD authors face limited citation reward or penalty by working on these geographically distinctive papers in part (d) of Figure 10. The outcome variable is measured as the number of citations garnered by a paper within 5 years from its publi-

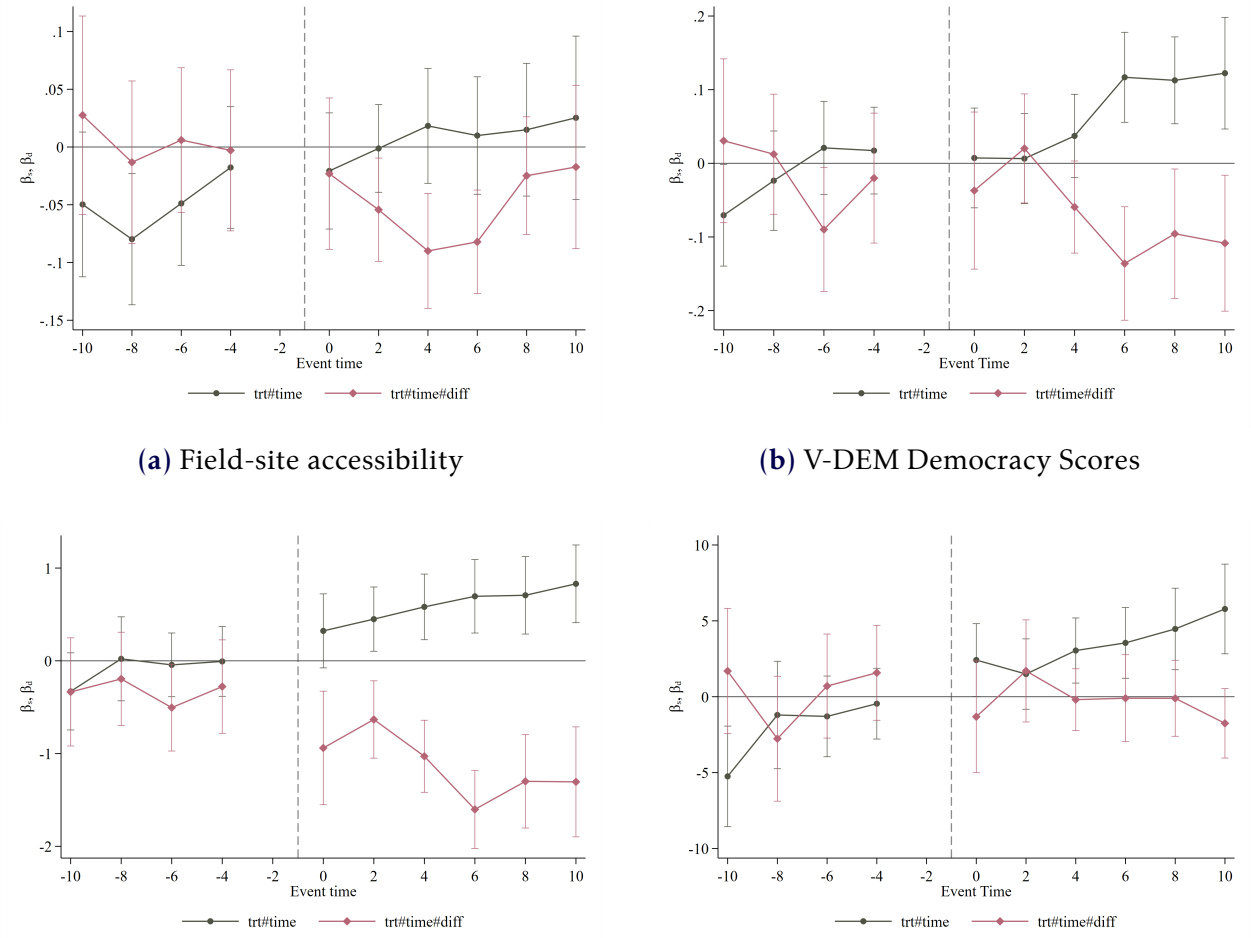
cation. In addition to the battery of controls used in computing coefficients for previous figures, here I add a control for benchmark citations, that is, the number of citations received by papers written by young researchers at the same university within five years of publication. Controls for journal ranking and year of publication ameliorate concerns that quality differences explain the lack of citation reward. Thus, despite facing additional barriers in producing research, LD authors do not receive concomitant awards. This provides indirect evidence of lack of incentives for pursuing such research in status quo and hints at the potential role of preferences of LD authors in making research-related choices. Thus, diversity-enhancing policies within scientific enterprise could deliver diverse research in the absence of additional incentives to promote such research.

6 Conclusion

Scientific research remains disproportionately concentrated on questions, populations, and environments relevant to high-income countries. Whether this pattern reflects the incentives embedded within scientific institutions or the composition of the scientific workforce itself has important implications for science policy. If incentives are the dominant force, then changing who enters science may have little effect on what science gets done. If researchers bring distinct information, experiences, and preferences into their work, then diversifying the scientific workforce may alter the direction of scientific attention.

This paper provides evidence for the latter channel. I exploit the staggered introduction of English-Taught Master's Programs at research-intensive universities in non-Anglophone countries and show that these programs substantially diversify the pool of early-career researchers entering scientific careers. Universities that attract researchers from more linguistically and geographically diverse backgrounds subsequently expand the set of countries, species, and environments they study. The increase is concentrated in research relevant to linguistically dissimilar countries and is especially pronounced in high-quality scientific outlets.

The results further suggest that these changes arise because incoming researchers possess knowledge, interests, and capabilities that differ from those of incumbent scientists. Researchers from linguistically dissimilar backgrounds disproportionately conduct research in more remote locations and in countries with weaker environmental and institutional outcomes, despite receiving little additional citation reward for doing so. These findings indicate that the composition of the scientific workforce can shape the allocation of research effort even when the incentive structure of science remains unchanged.



(a) Field-site accessibility

(b) V-DEM Democracy Scores

(c) Yale EPI Scores

(d) Citations

Notes: This figure plots coefficients from paper-level event-study-style regression which looks at characteristics of research sites of papers written by LS and LD authors. coefficients of interest - those describing characteristics of LD authors - are in pink. Measures used are road density (a), Democracy index (b), EPI scores (c) and citations received by the paper in first 5 years (d). 95% CIs are also shown. Unlike previous regressions, here two-year intervals are grouped together.

Figure 10: Differences in research pursued by LS and LD authors

7 Supplementary Tables

Table 3: Research groups within the Ecology and Conservation Science theme at the University of Cambridge Department of Zoology.

Lab / Group	Lead PI	Summary	Orientation	Taxon / Ecosystem	Distribution
Vertebrate Palaeontology Group	Jason J. Head	Studies deep-time evolution and biogeography of vertebrates using fossils and paleoecology.	Taxon-forward	Fossil vertebrates	Global / prehistoric
Marine Behavioural Ecology Group	James Herbert-Read	Examines collective movement, sociality, and predator–prey dynamics in marine organisms.	Question-forward	Marine fishes and aquatic animals	Global marine systems
Marine Ecology and Conservation Group	Catharine Horswill	Investigates marine ecosystem dynamics and conservation responses to environmental change.	Ecosystem-forward	Marine ecosystems	Global oceans
Insect Evolution and Genomics Group	Chris Jiggins	Uses genomics to study speciation, mimicry, and adaptation in insects.	Taxon-forward	*Heliconius* butterflies	~15–20 Latin American countries
Kilner Group	Rebecca Kilner	Studies social evolution, parental care, and family interactions in animals.	Question-forward	Burying beetles; *Nicrophorus*	~30–40 countries
LeBoeuf Lab / Social Fluids Lab	Adria LeBoeuf	Explores chemical communication and social physiology in insect societies.	Taxon-forward	Ants and social insects	>150 countries
Evolutionary Ecology Group	Andrea Manica	Studies how ecological pressures shape evolution, dispersal, and collective behavior.	Question-forward	Mixed taxa	N/A
Biotic Interactions Group / Miller Lab	Christine Miller	Investigates how species interactions drive ecological and evolutionary outcomes.	Question-forward	Mixed systems	N/A
Deep-time Ecology Group	Emily Mitchell	Reconstructs ancient ecosystems and macroecological patterns from the fossil record.	Question-forward	Ediacaran ecosystems	Multiple continents

Continued on next page

Lab / Group	Lead PI	Summary	Orientation	Taxon / Ecosystem	Distribution
Cichlid Eco-evo-devo Group	Emília Santos	Uses cichlid fishes to study links between development, ecology, and evolution.	Taxon-forward	Cichlid fishes	~40–50 countries
Conservation Science Group	William Sutherland	Develops evidence-based approaches for biodiversity conservation and policy design.	Question-forward	Conservation systems broadly	N/A
Insect Ecology Group	Edgar Turner	Studies tropical insect biodiversity, ecosystem functioning, and land-use change.	Taxon-forward	Tropical insects	Global; >190 countries

Table 4: Conservation and biodiversity research groups at Princeton EEB.

Group / Lab	Lead PI	Summary	Orientation	Taxon / Ecosystem	Distribution
Dobson Lab	Andrew Dobson	Studies infectious-disease ecology, host–pathogen dynamics, and conservation in fragile ecosystems.	Question/theme-forward	Pathogens, hosts, endangered species	N/A
Hedin Lab	Lars Hedin	Studies ecosystem analysis, nutrient cycling, greenhouse gases, and forest biogeochemistry across scales.	Ecosystem/theme-forward	Temperate and tropical forests	Global forest systems
Center for BioComplexity / Levin Group	Simon Levin	Develops theory on ecosystem structure, disease dynamics, and coupled ecological–socioeconomic systems.	Question/method-forward	Ecosystems broadly	N/A
Pacala Group	Stephen Pacala	Builds mathematical and ecological models of communities, ecosystems, climate feedbacks, and the global carbon cycle.	Question/method-forward	Vegetation, biomes, carbon cycle	Global
Pringle Lab	Robert Pringle	Studies how wild ecosystems work, especially food webs, biodiversity, restoration, and species interactions.	Ecosystem-forward	African savannas; Caribbean islands	Regional / multi-country
Rubenstein Group	Daniel Rubenstein	Studies animal decision-making, social behavior, movement, and conservation, especially in equids.	Taxon-forward	Equids: horses, zebras, asses	~40–50 countries
The Drongos	David Wilcove	Studies how to protect biodiversity in a hotter, more crowded world using ecology, conservation biology, and policy.	Question/theme-forward	Birds plus reptiles, mammals, fish, coral, plants, insects	Global; all continents except Antarctica

Table 5: First-stage effects by career stage

Coefficient	Early-Career	Mid-Career	Late-Career
$\beta_{(-10,-6)}$	0.001	-0.006	-0.010
	(0.002)	(0.004)	(0.005)
$\beta_{(-5,-1)}$	0.001	-0.002	-0.007
	(0.002)	(0.004)	(0.004)
$\beta_{(0,4)}$	0.018	0.013	-0.015
	(0.008)	(0.014)	(0.013)
$\beta_{(5,10)}$	0.061	0.021	-0.010
	(0.011)	(0.019)	(0.016)
ATT	0.051	0.029	-0.016
	(0.010)	(0.020)	(0.019)
Mean outcome	0.245	0.335	0.385
N	6,582	6,358	6,320
R-squared	0.713	0.694	0.749

Notes: Coefficient estimates are reported as entries, with standard errors in brackets below. I provide coefficients as pooled ATTs for roughly 5-year periods. The outcome variable is share of incoming LD researchers by career stage. Results for different career stages are reported separately.

Table 6: First-stage effects on distinct new nationalities

Coefficient	Early-Career	Mid-Career	Late-Career
$\beta_{(-10,-6)}$	-0.041	0.060	-0.106
	(0.088)	(0.080)	(0.073)
$\beta_{(-5,-1)}$	0.176	-0.021	0.089
	(0.110)	(0.094)	(0.089)
$\beta_{(0,4)}$	0.376	-0.136	-0.439
	(0.375)	(0.335)	(0.249)
$\beta_{(5,10)}$	1.935	0.687	-0.154
	(0.689)	(0.587)	(0.413)
ATT	2.098	0.617	-0.319
	(0.666)	(0.492)	(0.348)
Mean outcome	11.368	7.392	6.466
N	6,638	6,503	6,448
R-squared	0.841	0.818	0.778

Notes: Coefficient estimates are reported as entries, with standard errors in brackets below. I provide coefficients as pooled ATTs for roughly 5-year periods. The outcome variable is the number of different nationalities represented in incoming cohorts of researchers, by career stages.

Table 7: Event-study effects by linguistic similarity

	LS Papers		LD papers	
Coefficient	(a)	(b)	(a)	(b)
$\beta_{(-10,-6)}$	1.017	0.558	1.258	0.466
	(0.508)	(0.259)	(0.418)	(0.147)
$\beta_{(-5,-1)}$	0.913	0.502	1.220	0.690
	(0.469)	(0.262)	(0.393)	(0.131)
$\beta_{(0,4)}$	2.746	0.104	5.565	1.800
	(3.524)	(1.335)	(3.007)	(0.713)
$\beta_{(5,10)}$	10.912	2.231	18.745	7.835
	(9.284)	(2.958)	(8.914)	(2.025)
ATT	9.912	2.103	18.914	8.855
	(3.853)	(1.605)	(4.220)	(1.686)
Mean outcome	24.806	12.106	28.426	8.327
N	6,707	6,707	6,707	6,707
R-squared	0.700	0.747	0.693	0.703

Notes: Coefficient estimates are reported as entries, with standard errors in brackets below. I provide coefficients as pooled ATTs for roughly 5-year periods. With each category (LS/LD), (a) refers to estimates for papers that are geographically relevant to LS/LD countries, whereas (b) counts only those papers whose field site was recorded in LS or LD country.

References

Acemoglu, Daron and Fabrizio Zilibotti (2001) "Productivity differences," *The Quarterly Journal of Economics*, 116 (2), 563–606.

Ager, Philipp and Markus Brückner (2013) "Cultural diversity and economic growth: Evidence from the US during the age of mass migration," *European Economic Review*, 64, 76–97.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas (2017) "Immigration and the rise of American ingenuity," *American Economic Review*, 107 (5), 327–331.

Akerman, Ariel, Jacob Moscona, Heitor S Pellegrina, and Karthik Sastry (2025) "Public R and D Meets Economic Development: Embrapa and Brazil's Agricultural Revolution,"Technical report, National Bureau of Economic Research.

Albuquerque, Fábio and Paul Beier (2016) "Predicted rarity-weighted richness, a new tool to prioritize sites for species representation," *Ecology and Evolution*, 6 (22), 8107–8114.

Amano, Tatsuya and William J Sutherland (2013) "Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security," *Proceedings of the Royal Society B: Biological Sciences*, 280 (1756).

Andrews, Michael J and Alexa Smith (2025) "Do local conditions determine the direction of science? Evidence from US land grant colleges," *Explorations in Economic History*, 97, 101669.

Arkolakis, Costas, Sun Kyoung Lee, and Michael Peters (2020) "European immigrants and the United States' rise to the technological frontier," in *2019 Meeting Papers*, 1420, 61.

Arts, Sam, Nicola Melluso, and Reinhilde Veugelers (2025) "Beyond citations: Measuring novel scientific ideas and their impact in publication text," *Review of Economics and Statistics*, 1–33.

Ash, Elliott, David Cai, Mirko Draca, and Shaoyu Liu (2022) "Bootstrapping science? The impact of a "return human capital" programme on Chinese research productivity."

Association, Canadian Medical et al. (2004) "Western medical journals and the 10/90 problem."

Avdeev, Stanislav (2025) "University as a Melting Pot: Long-term Effects of Internationalization."

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin (2019) "Does science advance one funeral at a time?" *American Economic Review*, 109 (8), 2889–2920.

Bayer, Amanda and Cecilia Elena Rouse (2016) "Diversity in the economics profession: A new attack on an old problem," *Journal of Economic Perspectives*, 30 (4), 221–242.

Becker, Sascha O, Volker Lindenthal, Sharun W Mukand, and Fabian Waldinger (2024) "Persecution and escape: professional networks and high-skilled emigration from nazi Germany," *American Economic Journal: Applied Economics*, 16 (3), 1–43.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019) "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 134 (2), 647–713.

Block, Sebastián, John W Emerson, Daniel C Esty, Alex de Sherbinin, and Zachary A Wendling (2024) "2024 Environmental performance index."

Callaghan, Max, Carl-Friedrich Schleussner, Shruti Nath, Quentin Lejeune, Thomas R Knutson, Markus Reichstein, Gerrit Hansen, Emily Theokritoff, Marina Andrijevic, Robert J Brecha et al. (2021) "Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies," *Nature climate change*, 11 (11), 966–972.

Callaway, Brantly and Pedro HC Sant'Anna (2021) "Difference-in-differences with multiple time periods," *Journal of econometrics*, 225 (2), 200–230.

Cook, Lisa D (2020) "Policies to broaden participation in the innovation process," *The Hamilton Project*, 11, 1–23.

Das, Jishnu, Quy-Toan Do, Karen Shaines, and Sowmya Srikant (2013) "US and them: The geography of academic research," *Journal of Development Economics*, 105, 112–130.

De Souza, Gustavo (2022) "The Labor Market Consequences of Appropriate Technology,"Technical report, Federal Reserve Bank of Chicago.

Dickinson, Janis L, Benjamin Zuckerberg, and David N Bonter (2010) "Citizen science as an ecological research tool: challenges and benefits," *Annual review of ecology, evolution, and systematics*, 41 (1), 149–172.

Dimitris, Michelle C, Matthew Gittings, and Nicholas B King (2021) "How global is global health research? A large-scale analysis of trends in authorship," *BMJ global health*, 6 (1).

Dossi, Gaia (2024) "Race and science,"Technical report, Working Paper.

Einiö, Elias, Josh Feng, and Xavier Jaravel (2019) "Social push and the direction of innovation," *American Economic Review*.

Fazey, Ioan, Joern Fischer, and David B Lindenmayer (2005a) "What do conservation biologists publish?" *Biological conservation*, 124 (1), 63–73.

——— (2005b) "Who does all the research in conservation biology?" *Biodiversity & Conservation*, 14 (4), 917–934.

Fick, Stephen E and Robert J Hijmans (2017) "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas," *International journal of climatology*, 37 (12), 4302–4315.

Fry, Caroline and Megan MacGarvie (2024) "Author country of origin and attention on open science platforms: Evidence from COVID-19 preprints," *Management Science*, 70 (8), 5426–5444.

Ganguli, Ina (2015) "Immigration and ideas: what did Russian scientists "bring" to the United States?" *Journal of Labor Economics*, 33 (S1), S257–S288.

Geldmann, Jonas, Jacob Heilmann-Clausen, Thomas E Holm, Irina Levinsky, BO Markussen, Kent Olsen, Carsten Rahbek, and Anders P Tøttrup (2016) "What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements," *Diversity and Distributions*, 22 (11), 1139–1149.

Govaerts, Rafaël, Eimear Nic Lughadha, Nicholas Black, Robert Turner, and Alan Paton (2021) "The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity," *Scientific data*, 8 (1), 215.

Henrich, Joseph, Steven J Heine, and Ara Norenzayan (2010) "The weirdest people in the world?" *Behavioral and brain sciences*, 33 (2-3), 61–83.

Hill, Ryan and Carolyn Stein (2025a) "Race to the bottom: Competition and quality in science," *The Quarterly Journal of Economics*, 140 (2), 1111–1185.

——— (2025b) "Scooped! Estimating rewards for priority in science," *Journal of Political Economy*, 133 (3), 793–845.

Hughes, Alice C, Michael C Orr, Keping Ma, Mark J Costello, John Waller, Pieter Provoost, Qinmin Yang, Chaodong Zhu, and Huijie Qiao (2021) "Sampling biases shape our view of the natural world," *Ecography*, 44 (9), 1259–1269.

Hunter, Rosalind S, Andrew J Oswald, and Bruce G Charlton (2009) "The elite brain drain," *The Economic Journal*, 119 (538), F231–F251.

Jambeck, Jenna R, Roland Geyer, Chris Wilcox, Theodore R Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law (2015) "Plastic waste inputs from land into the ocean," *science*, 347 (6223), 768–771.

Kadmon, Ronen, Oren Farber, and Avinoam Danin (2004) "Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models," *Ecological Applications*, 14 (2), 401–413.

Kantor, Shawn and Alexander Whalley (2019) "Research proximity and productivity: long-term evidence from agriculture," *Journal of Political Economy*, 127 (2), 819–854.

Keding, Lisa and Marten C Ritterrath (2025) "Inventors' personal experience of natural disasters and green innovation,"Technical report, ECONtribute Discussion Paper.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson (2021) "Who do we invent for? Patents by women focus more on women's health, but few women get to invent," *Science*, 372 (6548), 1345–1348.

Ku, Hyejin and Asaf Zussman (2010) "Lingua franca: The role of English in international trade," *Journal of Economic Behavior & Organization*, 75 (2), 250–260.

Linxen, Sebastian, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke (2021) "How weird is chi?" in *Proceedings of the 2021 chi conference on human factors in computing systems*, 1–14.

Lundberg, Shelly and Jenna Stearns (2019) "Women in economics: Stalled progress," *Journal of Economic Perspectives*, 33 (1), 3–22.

Melitz, Jacques (2008) "Language and foreign trade," *European Economic Review*, 52 (4), 667–699.

Melles, Stephanie J, Christopher Scarpone, Adisa Julien, Julie Robertson, Jennifer Bello Levieva, Courtney Carrier, Robert France, Sabrina Guvenc, Wai Ying Lam, Michelle Lucas et al. (2019) "Diversity of practitioners publishing in five leading international journals of applied ecology and conservation biology, 1987–2015 relative to global biodiversity hotspots," *Ecoscience*, 26 (4), 323–340.

Moscona, Jacob and Karthik A Sastry (2025) "Inappropriate technology: Evidence from global agriculture,"Technical report, National Bureau of Economic Research.

Moser, Petra, Alessandra Voena, and Fabian Waldinger (2014) "German Jewish émigrés and US invention," *American Economic Review*, 104 (10), 3222–3255.

Nagaraj, Abhishek and Randol Yao (2026) "The Geography of Science,"Technical report, National Bureau of Economic Research.

Nocito, Samuel (2021) "The effect of a university degree in English on international labor mobility," *Labour Economics*, 68, 101943.

Parey, Matthias, Jens Ruhose, Fabian Waldinger, and Nicolai Netz (2017) "The selection of high-skilled emigrants," *Review of Economics and Statistics*, 99 (5), 776–792.

Parey, Matthias and Fabian Waldinger (2011) "Studying abroad and the effect on international labour market mobility: Evidence from the introduction of ERASMUS," *The economic journal*, 121 (551), 194–222.

Pitt, Catherine, Catherine Goodman, and Kara Hanson (2016) "Economic evaluation in global perspective: a bibliometric analysis of the recent literature," *Health Economics*, 25, 9–28.

Posch, Max, Jonathan Schulz, and Joseph Henrich (2026) "How cultural diversity drives innovation: Surnames and patents in us history," *Journal of Political Economy*, 134 (4), 1245–1305.

Rambachan, Ashesh and Jonathan Roth (2023) "A more credible approach to parallel trends," *Review of Economic Studies*, 90 (5), 2555–2591.

Roberts, Beth EI, W Edwin Harris, Geoff M Hilton, and Stuart J Marsden (2016) "Taxonomic and geographic bias in conservation biology research: a systematic review of wildfowl demography studies," *PloS one*, 11 (5), e0153908.

Schmallenbach, Leo, Maximilian Bley, Till W Bärnighausen, Cassidy R Sugimoto, Carolin Lerchenmüller, and Marc J Lerchenmueller (2025) "Global distribution of research efforts, disease burden, and impact of US public funding withdrawal," *Nature Medicine*, 31 (9), 3101–3109.

Shattuck, Annie, Marion Werner, Finn Mempel, Zackary Dunivin, and Ryan Galt (2023) "Global pesticide use and trade database (GloPUT): New estimates show pesticide use trends in low-income countries substantially underestimated," *Global Environmental Change*, 81, 102693.

Stansbury, Anna and Robert Schultz (2023) "The economics profession's socioeconomic diversity problem," *Journal of Economic Perspectives*, 37 (4), 207–230.

Subramani, Gauri and Caroline Fry (2025) "Migration and Local Problem-Solving," in *Academy of Management Proceedings*, 2025, 11212, Academy of Management Valhalla, NY 10595.

Suri, Tavneet and Christopher Udry (2022) "Agricultural technology in Africa," *Journal of Economic Perspectives*, 36 (1), 33–56.

Truffa, Francesca and Ashley Wong (2025) "Undergraduate Gender Diversity and the Direction of Scientific Research," *American Economic Review*, 115 (7), 2414–2448.

Wilson, Kerrie A, Nancy A Auerbach, Katerina Sam, Ariana G Magini, Alexander St L Moss, Simone D Langhans, Sugeng Budiharta, Dilva Terzano, and Erik Meijaard (2016) "Conservation research is not happening where it is most needed," *PLoS Biology*, 14 (3), e1002413.