Colonial Knowledge Legacies and Scientific Recognition: A Data-Geographic Analysis of Epistemic Inequality in Global Prize Distribution

---

Introduction

Scientific recognition is one of the most powerful forms of symbolic capital in the global knowledge economy. Prestigious awards such as the Nobel Prize, the Fields Medal, and numerous national and disciplinary honors not only signal individual excellence but also shape institutional reputations, funding flows, and research priorities. Despite the growing internationalization of research networks and publishing infrastructures, the recipients of these honors remain disproportionately concentrated in a small number of institutions and countries — primarily in Western Europe and North America. This persistent concentration raises critical questions about the historical and structural forces that govern access to scientific prestige.

This study proposes a formal investigation into the role of colonial histories in shaping the contemporary geography of scientific recognition. While existing literature on global science inequality has focused on present-day disparities in resources, infrastructure, and research output, far less attention has been paid to how colonial knowledge regimes have contributed to durable epistemic hierarchies. By integrating postcolonial theory, the history of science, and computational data geography, we aim to assess whether the spatial distribution of scientific prizes reflects path-dependent institutional advantages rooted in colonial-era academic and intellectual networks.

The central hypothesis of this research is that colonial entanglements — including the establishment of scientific institutions by imperial powers, the extraction of intellectual labor, and the imposition of Western epistemological norms — have created lasting asymmetries in the distribution of symbolic scientific capital. Countries that were former colonizers may not only have greater representation among prize recipients, but also more influence over prize-giving institutions and the epistemic standards by which excellence is judged. Conversely, postcolonial nations may remain under-recognized despite growing contributions to global research output.

This project seeks to operationalize these hypotheses by constructing a longitudinal, geocoded dataset of scientific prizes and linking it to colonial history records and contemporary research metrics. By doing so, it becomes possible to quantitatively assess the relationship between historical colonial dynamics and current patterns of scientific recognition, offering a rigorous empirical contribution to debates about global knowledge equity.

Methods

Data Compilation and Historical Annotation

The first stage involves constructing a comprehensive database of internationally recognized scientific prizes from 1900 to 2023. This includes major awards such as the Nobel Prizes, Lasker Awards, Fields Medal, Breakthrough Prizes, and other globally or regionally significant honors across the natural sciences, medicine, social sciences, and engineering. For each prize, we record the year of award, name of recipient, nationality, institutional affiliation at the time of the award, field of research, and location of the awarding body. Institutional data is geocoded using standardized national boundaries and linked to country-level identifiers.

To account for the historical dimension of colonialism, we integrate structured datasets detailing colonial relationships, including colonizer-colonized dyads, duration of colonization, year of independence, type of colonial administration (e.g., settler, extractive, mandate), and institutional legacies such as transplanted education systems or scientific academies. Each modern-day country is assigned a composite *colonial legacy score*, derived from (1) the length and intensity of colonial rule, (2) linguistic and institutional continuity with the former colonizer, and (3) the presence of historically significant colonial scientific infrastructure (e.g., botanical gardens, observatories, universities).

We complement this with current scientific indicators — national R&D expenditure, total publication output, citation metrics, and number of active researchers — to control for contemporary capacity. All data is harmonized into a panel structure, enabling country-level and institutional-level analysis over time.

Statistical Modeling and Spatial Analysis

The primary analytic goal is to model the relationship between colonial legacy and scientific recognition outcomes. The dependent variable is defined as the number of international scientific prizes awarded to individuals affiliated with institutions in a given country, normalized by population and research output. Independent variables include the colonial legacy score, GDP per capita, number of top-tier research institutions, language alignment with prize-granting bodies, and regional location.

We estimate multiple regression models — including negative binomial models for count data and fixed-effects panel regressions — to isolate the effect of colonial legacy while accounting for temporal and regional variation. Matching techniques are used to compare countries with similar present-day research outputs but different colonial histories, allowing for quasi-experimental inference about the influence of historical structures. We also stratify the analysis by discipline and prize origin to identify whether colonial effects are more pronounced in certain fields or institutions.

To capture spatial dependencies, we perform spatial autocorrelation analysis (Moran's I) and geospatial clustering to assess the degree of geographic concentration in prize distribution.

These spatial statistics are used to map persistent "epistemic cores" and "peripheral zones" of recognition, and to track how these patterns evolve over the past century. Additionally, we conduct a network analysis of institutional affiliations and training lineages among prize recipients, highlighting whether scientific prestige tends to circulate within historically privileged institutional networks with colonial ties.

All models include robustness checks for multicollinearity, omitted variable bias, and temporal autocorrelation. Data and code will be made openly available to facilitate replication and further investigation.

Potential Impact

This research has the potential to reshape how we understand scientific recognition, not as a neutral outcome of intellectual merit but as a product of historical and geopolitical structures that continue to shape the distribution of symbolic capital. By demonstrating a quantifiable link between colonial histories and present-day recognition patterns, we provide empirical grounding to theoretical claims from postcolonial and critical science studies about the reproduction of epistemic inequality in the global scientific system.

The findings can contribute to policy debates on international science funding, institutional reform, and diversity in recognition bodies. Prize committees and academic institutions seeking to decolonize their evaluative practices may benefit from concrete data on geographic and institutional biases, enabling more reflective and inclusive criteria for recognition. Likewise, science diplomacy efforts aimed at fostering global equity may use these insights to address structural imbalances in symbolic authority and legitimacy.

More broadly, this project highlights the importance of historical data in computational social science. By combining bibliometric techniques with colonial history and data geography, we demonstrate that the spatial dynamics of prestige in science cannot be fully understood without accounting for the temporal depth and political contours of global knowledge systems. In doing so, the study opens new avenues for interdisciplinary research at the intersection of history, data science, and epistemology, and challenges the prevailing narrative that global scientific excellence is evenly distributed or naturally meritocratic.