THE UNIVERSITY OF CHICAGO

THE STRUCTURE OF KNOWLEDGE DIFFUSION IN SCIENCES AND CONSEQUENCES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF SOCIOLOGY

BY

DONGHYUN KANG

CHICAGO, ILLINOIS

AUGUST 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to my advisor, James Evans, whose encouragement, understanding, patience, optimism, and intellectual stimulation have greatly enriched my Ph.D. journey at the University of Chicago. His unwavering support during the challenging times of my research has helped me to persevere and succeed. I am immensely grateful for his guidance and steadfast belief in my potential and our projects. I am also sincerely thankful to my dissertation committee members, John Levi Martin and Karin Knorr-Cetina. John Levi Martin's guidance and intellectual rigor, evident in his writings and through his classes and discussions from the first year of my graduate studies, have profoundly shaped my attitude toward sociological science. I also thank Karin Knorr-Cetina, whose involvement has profoundly impacted my intellectual perspective. Moreover, her insightful suggestions for readings in science studies have been truly invaluable. Their engagement has been crucial to my academic development.

My sincere thanks also go to the members of the Knowledge Lab for providing the resources and an ideal, vibrant environment for computational social science. I must recognize their support and the thoughtful conversations over the years, including but not limited to both current and past lab members: Candice Lewis, Ziwen Chen, Di Tong, Jeremiah Milbauer, Alexander Belikov, Brandan Chambers, Jamshid Sourati, Lingfei Wu, Bhargav Srinivasa Desikan, Deblina Mukherjee, Clara del Junco, Fengli Xu, Likun Cao, Xin Gao, Haochuan Cui, Chris Esposito, Stefan de Jong, Eamon Duede, Nak Won Rim, Austin Kozlowski, Jacy Reese Anthis, Junsol Kim, Jeff Lockhart, Hongkai Mao, Hyunku Kwon, and Haizi Yu. All the conversations, jokes, and intellectual engagements have enriched my academic journey. I look forward to even greater times in the near and far future.

I appreciate my collaborators for past and ongoing research projects, including TaeYoung Kang, Junkyu Jang, Renli Wu, Simon Yamawaki Shachter, Wenxuan Shi, Carina Kane, and Shiyang Lai. I appreciate their patience, time, and efforts. I sincerely hope our collaboration leads to fruitful outcomes and continues to grow. I also thank the members of my writing group: Andrew Swift, Xiangyu Ma, Tim Elder, and Hong Jin Jo. I have completed a plethora of writings during our collective writing sessions, which I do not think I could have accomplished without this shared working setting. I must also recognize the 2017 Chicago Sociology Ph.D. cohort members (who have not been mentioned above): Anna Berg, Ilana Ventura, Kailey White, Ariel Azar, Stephanie Ternullo, Yuchen Yang, Maurice Bokanga, Teng Ge, Nisarg Mehta, which I consider the best of the best.

I must extend my deepest gratitude to my parents, whose steady support and understanding have been fundamental throughout this journey. Their encouragement has been my anchor during my academic pursuits' inevitable highs and lows. Lastly, my most heartfelt thanks should go to So Yun Kim, my darling wife, whose love, and companionship are more precious to me than any words can convey. Her presence has brought a profound sense of purpose to my daily endeavors. She has been the cornerstone of my daily life, infusing moments with purpose. I am grateful for her and our son, Lucas, who brings previously unfathomable happiness and love into my life. I dedicate this dissertation to them.

ABSTRACT

The enterprise of modern science evolves alongside its underlying structure, encompassing interactions among scientists, institutions, and culture, as well as the development of scientific ideas and discoveries. The increasing availability of digitized scholarly data, coupled with enhanced computational power, offers unprecedented opportunities to characterize, critically examine, and untangle how these complex, intertwined entities shape the inner workings of science. Through three empirical studies, this dissertation investigates how underlying social and infrastructural elements can influence the production, diffusion, and consumption of scientific ideas and discoveries. The introduction outlines the three studies. The first chapter tackles the question of how tacitly encoded configurations and undocumented components can influence the estimates from randomized clinical trials, demonstrating the role of socio-epistemic bubbles in the production of scientific medical knowledge. The results suggest why meta-analyses may not mechanically resolve scientific disagreements and disputes, contrary to widespread expectations. The second chapter extends the analogy of bubbles to the realm of attention in biomedical scientific knowledge, reflecting the phenomenon of bubbles and collapses in financial asset markets. It reveals that restricted diffusion within social and scientific 'bubbles' can precede sudden collapses in scientific attention, offering a straightforward framework for identifying early signs of these bubbles in science. The third study explores how the presence of code repositories alongside machine learning research affects the citation rates of papers. It finds that the popularity of ML frameworks, such as PyTorch and TensorFlow, used in these repositories can have second-order network effects, underscoring the latent role of technological artifacts and infrastructure in scientific dissemination. The final chapter concludes the dissertation with reflections and outlines future research directions.

Introduction

As a social institution, the modern enterprise of science is shaped not only by scientific ideas and discoveries but also by interactions among scientists, institutions, research infrastructure, and diverse research cultures. This complex interconnectedness has motivated systematic inquiries to understand operations, the social conditions underpinning them, and the individuals involved in sciences (Collins 1983; Ben-David and Sullivan 1975; Kim 1994; Bourdieu 2004). The abundance of digitized scholarly data, coupled with enhanced computational capabilities, offers unprecedented opportunities to characterize, analyze, and untangle how these intertwined associations of entities, leading to an emerging multidisciplinary field of Science of Science (Fortunato et al. 2018; Wang and Barabási 2021). The following chapters in the dissertation, which collectively explore the underlying latent structures of scientific communities and their impacts, aim to bridge computational large-scale analysis with theoretical insights from the sociology of knowledge and science to study the production, consumption, and diffusion of scientific knowledge at scale.

Chapter 1 conceptualizes science as a field of knowledge characterized by overlapping socio-epistemic bubbles, within which scientific practices, tacit, knowledge, and implicit preferences, and heuristics are shaped and circulated by localized networks of scientists and researchers (Collins 2010; Knorr-Cetina 1999; Crane 1972). Doing so leads to a new theorization that tacit knowledge and scientific practices, typically seen as transferred through direct human interaction or embodied experiences, occur and reside across an epistemic field (Martin 2003; Bourdieu 2004). These socio-epistemic bubbles are socially constituted and reinforced (Knorr-Cetina 1999), localized rather than universal, and can remain invisible (de Solla Price and Beaver 1966; Crane 1972) to those within and outside them. This theorization leads to a hypothesis that

even estimates drawn from randomized clinical trials may also be influenced by these socio-epistemic bubbles.

To empirically test the influence of these socio-epistemic bubbles, the first chapter examines randomized clinical trials (RCTs) collected by Cochrane Systematic Reviews, one of the most authoritative sources of evidence-based medicine practices. To measure the socio-epistemic fields and bubbles, a computational manifold embedding technique (Le and Mikolov 2014) is adapted to inscribe relations among collaborating researchers and articles produced through co-authorship, trained on approximately 8.4 million disambiguated authors and 28.3 million articles from Pubmed Knowledge Graph (Xu et al. 2020). Consequently, the embedding space produces similar vector positions for authors who frequently co-author papers and papers co-authored by frequent collaborators. Then, clinical trials curated within 1,962 Cochrane Systematic Reviews were projected to this measurable field of social relationships and scientific knowledge. The analysis reveals that clinical trials closer to each other within this social space exhibit a higher degree of homogeneity. Furthermore, I found that statistically significant results are more likely to be challenged by distant researchers, while early null findings tend to gain statistical support when later studied by researchers close to the initial study authors. These results underscore the role of collaborative relationships as a subtle but critical proxy for capturing socio-epistemic bubbles, unlabeled but latent sources of variation in RCTs that affect the production side of scientific biomedical knowledge.

Chapter 2 draws on the analogy between bubbles in asset markets and the potential inflation of attention in the sciences, further extending the concept of social and epistemic bubbles. This conceptualization builds upon previous considerations from philosophers and observers of science, who suggest that agents in science can be viewed as vendors selling their

research (Goldman and Shaked 1991) and that the scientific system as a whole is a social investment not immune to the phenomenon of bubbles—often recognized in financial markets (Pedersen and Hendricks 2014; Evans et al. 2011). Combining this with the widely held view that citation counts serve as a measure of the importance and impact of scientific work despite imperfection (Fortunato et al. 2018; Partha and David 1994), the chapter seeks to identify a leading signal associated with an abrupt drop in scientific attention towards seemingly promising subfields in biomedical science. Inspired by the case of cardiac regeneration research in biomedicine, it posits that a lack of genuine diffusion, not captured by citation counts—or indicative of scientific bubbles—may predict a rapid decline in popularity or the burst of scientific bubbles.

Building on this, I introduce the 'diffusion index,' designed to quantify the degree of research diffusion across intellectual and social spaces. This was achieved by assigning distances within citation networks based on two embedding models: 1) the 'social space' that maps the collaboration network of biomedical scientists, and 2) the 'scientific space' that encodes the direct and indirect associations of Medical Subject Headings. Applying this framework to 28,504 unique biomedical subfields (Azoulay, Fons-Rosen, and Zivin 2019), the analysis suggests that limited diffusion precedes a drastic decline in a given subfield's popularity. This result indicates that restricted diffusion in science can effectively identify scientific bubbles and predict their collapse. Additional analyses explore the consequences of these collapses, such as the association between bubbles and the continuous funding of irrelevant biomedical projects or the impact of entering a field near a bubble's burst on the long-term reputational damage to a scientific career. These findings further highlight the significance of findings and the importance of early detection of potential bubbles.

Chapter 3 turns the attention to the domain of machine learning (ML) research and also another critical aspect of contemporary scientific endeavor: the code repositories associated with research papers. In recent years, scientists and researchers have been encouraged, expected, or often required to share scientific artifacts along with their manuscripts—such as data and model implementations—via publicly accessible platforms. This has prompted a strand of scholarship that investigates the incentives, motivations, and costs linked with research transparency (Kim and Adler 2015; Mukherjee and Stern 2009; Wilms et al. 2020), as well as the impact of data and material sharing on the citation trajectories of research articles and broader implications (Kwon and Motohashi 2021; Furman and Stern 2011; Christensen et al. 2019). Against this backdrop, Chapter 3 aims to evaluate the degree to which the availability of code repositories, particularly GitHub repositories associated with papers, affects the citation rates of ML research articles. It also examines the impact of the choice of ML framework, such as PyTorch and TensorFlow, on the citations of research papers.

To examine this, I established a comprehensive linkage between article records cataloged in Papers with Code (PwC), the largest platform through which ML research articles and codes are linked, and Microsoft Academic Graph (MAG). Then, utilizing a random sample of approximately 20,000 ML articles linked to GitHub repositories along with papers topically similar but without repositories, the first analysis demonstrates that papers with repositories, on average, have about 20% advantages in monthly citation rates after the creation of the first GitHub repositories. Subsequently, the second analysis shows that the popularity trends of ML frameworks influence the monthly citation rate of related ML papers, exerting second-order network effects (Economides and Salop 1992). Together, these findings suggest the importance of technological infrastructure and artifacts in the diffusion of research.

The Conclusion closes the dissertation with summaries and reflections on each chapter. I also discuss the stakes of computational analysis on the scientific system and outline some avenues for future research by highlighting potential affordance and challenges.

Chapter 1

Socio-Epistemic Bubbles and Tacit Confidence in Randomized Clinical Trials*

Abstract

The paradigm of scientific medicine is among the most influential epistemic shifts in the past century, wherein randomized clinical trials (RCTs) represent the impartial arbiter of legitimate medical knowledge, a view prevalent among quantitative social scientists. Nevertheless, not all RCTs agree, and systematic reviews are invoked to reconcile them. These assume the wisdom of crowds, which hinges on diverse perspectives and data, across the distribution of analyzed studies, but socio-epistemic bubbles across them may reduce realized diversity. We theorize how tacit knowledge, beliefs, and expectations accumulate within these 'socio-epistemic bubbles,' continuous regions of latent social density that may decrease diversity and increase certainty about healthcare studied by RCTs. To assess our theory, we analyze the Cochrane systematic review repository, covering 20,117 meta-analyses extracted from 1,962 reviews. We find that being closer within 'social space' inscribed by scientific collaboration markedly increases agreement regarding RCT effect direction and size. Our analysis suggests that this amplified certainty can drive premature convergence and path-dependency affecting medical practice and population health. Moreover, our findings imply hidden limitations associated with unmeasured social influence across the policy sciences through which conflicting claims perpetuate and highlight the necessity of accounting for them to improve collective certainty

* Co-authored with James A. Evans, Department of Sociology, University of Chicago and Santa Fe Institute. I appreciate Daniel Yekutieli and Ruth Heller for discussions in the early stage of this work. I also thank the University of Chicago's Knowledge Lab members for valuable feedback. This work was completed in part with resources provided by the University of Chicago's Research Computing Center.

Introduction

Scientific advances and certainties represent a major contribution to modern economic growth and collective prosperity (Jones and Summers 2020; Oreskes 2019). The largest scale of scientific investments with the most immediate impact on human health and happiness lies in biomedicine (Ahmadpoor and Jones 2017). In the 21st Century, however, concerns have grown regarding the reproducibility and replicability[1] of published biomedical claims (Ioannidis 2005; Rzhetsky et al. 2006; Head et al. 2015; Krauss 2018), often characterized as the "reproducibility crisis" (Baker 2016). The first criticisms were raised about medical, pharmacological and genetic findings (Ioannidis 2005; Rzhetsky et al. 2006), but these soon spread to the behavioral sciences (Peterson and Panofsky 2021) and beyond (Baker 2016).

The "reproducibility crisis" may represent just another conflict in a long history of scientific disagreements between defensive and upstart ideas. Scientific development and change are rarely smooth, as rendered in Kuhn's well-known portrait of scientific revolutions (Collins 2000; Shwed and Bearman 2010; Kuhn 1962). What makes the contemporary concern over scientific replication distinctive is that it transpired despite seeming consensus over the method of generating rigorous evidence among the disciplines that employ statistical inference: randomized controlled trials (RCTs).

RCTs aim to minimize the impact of subjectivity by randomly assigning study subjects into control and treatment groups to offset confounding factors researchers cannot directly

[1] The 2019 report from the U.S. National Academies of Science, Engineering, and Medicine (NASEM) defines *reproducibility* as "obtaining consistent computational results using the same input data, computational steps, methods, and conditions of analysis," while *replicability* as "obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its data." (NASEM 2019: 1) Nevertheless, as hinted from a replication project name, *Reproducibility Project: Psychology* (Open Science Collaboration 2015*)* and the title of a manifesto, *A manifesto for reproducible science* (Munafò et al. 2017), two concepts have been often used interchangeably.

control. This allows scientists to achieve "mechanical objectivity" (Eyal 2019) based on explicit and transparent protocols. RCTs can be viewed as a vessel of Mertonian norms, especially that of universalism demanding that "pre-established impersonal criteria" (Merton 1973) be used to validate scientific claims. What is expected from this methodological apparatus is immunity from subjective bias, bursting the self-reinforcing filter bubbles and shattering the echo-chambers that persist in political and cultural domains (Bishop 2009; Pariser 2011; Sunstein 2018). Even though the RCT is not the only method of scientific inquiry, it occupies a pinnacle position in science, especially in the biomedical and policy domains where interventions are systematically evaluated for their effectiveness. It has notably risen as an arbiter of causal knowledge in the social and policy sciences in recent years, with five of the last eight Nobel awardees in Economics actively developing social scientific RCTs.

RCTs in reality, however, are more complex and uncertain than this aspiration allows. Consider the statistical properties of an RCT. A study can reject the null hypothesis or fail to do so simply by chance, a concern that looms large over studies of small sample sizes. A popular method to generalize across different RCTs is to conduct a meta-analysis, a secondary statistical evaluation pooling estimates from multiple primary studies.[2] Meta-analysis is the methodological centerpiece of a growing field called metascience, focused on robust, reproducible scientific knowledge (Munafò et al. 2017). Meta-analyses became widely adopted in scientific fields using RCTs, especially in medicine with the rise of evidence-based medicine (EBM) in the 1980s and early 1990s. It has garnered additional attention in the wake of the perceived replication crisis as a crucial tool to produce scientific synthesis with accumulating evidence (Munafò et al. 2017).

[2] In principle, non-RCT based studies can be included in a meta-analysis.

Meta-analysis relies upon an aggregation of individual studies analogous to "wisdom of crowds"—the phenomenon that aggregated judgments, decisions, and even speculations from a crowd of individuals give rise to better estimates than those made by a single one (Galton 1907).[3] This hinges on the independence and diversity of crowd members' information (Surowiecki 2004) and approach (Page 2019). In scientific crowds, findings established by more diverse researchers and distinct methods are much more likely to replicate (Danchev, Rzhetsky, and Evans 2019; Belikov, Rzhetsky, and Evans 2022). In a meta-analysis, each study is assumed to be drawn from the pool of studies on the same study topic using comparable research methods but independent of each other to offset idiosyncratic variations. Studies have demonstrated that social influence can cause herding in collective estimates (Lorenz et al. 2011; Da and Huang 2020), but subtle social and other factors have rarely been incorporated into meta-analyses.

Science studies scholars have long documented the difficulty of exact experimental replication, knowledge transfer, and communication (Collins 1985, 2001; Doing 2004). Tacit, uncodified knowledge is critical in experimental successes (Knorr-Cetina 1999) and transfers through social interaction (Collins 1985, 2010). A similar concern has played out in medicine, with many practitioners claiming that their subtle understanding of the "whole patient" results in a skilled art of diagnosis and treatment that cannot be reduced to the shallow equivalences of scientific medicine where medical subjects are treated interchangeably (Montgomery 2006). This concern inspires the research question we pose here: Are RCT estimates more similar among socially proximate scientists than distant ones?

Tacit knowledge has typically been discussed as only observable and transferable through direct human interaction or transferred as embodied experience. We broaden this phenomenon,

[3] Galton (1907) reported that the median value from the collective guessing of an ox's weight made by laypeople was almost close to actual weight.

theorizing scientific practices as occurring across an epistemic field (Martin 2003; Bourdieu 2004). Complex configurations of tacit expectations, often undocumented preferences, and differences in the logic of discovery and justification within a field form epistemic cultures (Knorr-Cetina 1999) that we call *socio-epistemic bubbles*. We append "socio" to "epistemic" because they are socially constituted and reinforced, and label them bubbles because they are often localized rather than universal and remain invisible to those inside and outside them. Scientific practices circulate within these bubbles through scientific habitus (Bourdieu 1975) to precisely shape bio-science experiments, beyond that which is documented in a research paper or protocol. In this paper, we demonstrate that such socio-epistemic fields and bubbles can be captured through a computational manifold embedding technique that inscribes relations among collaborating researchers manifest through co-authorship. Our evaluation demonstrates that RCTs conducted in healthcare settings are significantly and substantially more likely to report homogeneous estimates when within a socio-epistemic bubble—nearby in the measurable field of social relationships and scientific knowledge. These bubbles of agreement become reinforced by proximate insiders but can be burst by distant outsiders.

This paper extends insights from the social studies of science and connects to the growing literature on metascience by bridging the long-documented importance of tacit components in scientific knowledge production and concerns regarding replication in science. We extend methods at the intersection of network analysis and manifold learning to demonstrate the role of collaborative relationships as a subtle but critical proxy for capturing unlabeled but latent sources of variation in RCTs. We argue that this is why RCTs and meta-analyses often do not meet expectations as an impartial adjudicator. We further discuss the need to cultivate a more diverse community of researchers and scientists productively engaging with each other in order to

sustainably construct robust science with more widespread relevance (Danchev, Rzhetsky, and Evans 2019; Belikov, Rzhetsky, and Evans 2022).

Socio-Epistemic Fields and Bubbles

Tacit Knowledge and Replication in Science and Technology

Unlike explicit knowledge that can be codified in words or mathematical symbols, tacit knowledge is formed through experience, making it elusive to articulate.[4] Michael Polanyi famously illustrates its existence with bicycle riding: the rider cannot precisely articulate the process (Polanyi 1958); he"knows more than he can tell" (Polanyi and Sen 1966). Social studies of science have provided ample empirical evidence demonstrating the role of tacit knowledge in replication. Harry Collins notably showed that scientists could not build a TEA laser based on published information alone; essential information for the construction of the laser device traveled through migrating persons who had previously succeeded in building one (Collins 1985, 1974). Later ethnographies highlighted the role of embodied knowledge in a physics lab (Doing 2004), a molecular biology lab (Fujimura 1988), and a nuclear weapons program (MacKenzie and Spinardi 1995). Literature from science studies reveals that the complex nature of scientific inquiry often challenges clear and unambiguous forms of articulation, requiring interpersonal connections and immersive experience to render a discovery process observable and transferable.

Tacit knowledge has also been a central topic in innovation and knowledge management studies. This strand of work shows how unspoken, or undocumented, "know-how" is stored, practiced and diffused across individuals and organizations (Cowan and Foray 1997; Becker et

[4] Even codification can be shaped by socially guided implicit preferences and expectations. For instance, Warwick's (2011) study on the rise of mathematical physicists at Cambridge during the mid-19th to early-20th centuries offers a compelling illustration of how tacit knowledge and styles are passed down through teaching and academic training.

al. 2005). Some tacit knowledge, like embodied know-how, may elude straightforward codification. Other forms of tacit knowledge, while in principle explicable, can remain uncodified for diverse reasons. Things may become taken-for-granted, or their value may not merit the prohibitive cost of formalization (Nelson 2003). Such observations underscore the importance of shared tacit knowledge in replication rates across science and technology.

Reproduction is typically conceived as a successful attempt to produce the same answer by testing the same phenomena with the same methods on the same data. Replicability involves testing the same phenomena with the same methods and new data (NASEM 2019). Conceptual replication involves testing the same phenomena with different (while comparable) methods and different data, the hardest of the three (Belknap and Leonard 1991). Variation in reproduction or replication implies the existence of uncodified elements varying between them. The conditions of (conceptual) replication, by design, require tests conducted in different contexts, sometimes with different methods, such that they represent an attempt to vary uncodified elements from the original study, which were presumably non-critical to the claimed finding. The greater the distance between the original and the successfully replicated test, the more independent the evidence that replication provides. Deploying a new treatment regime from one patient, hospital, or country to another represents the critical (conceptual) replication hoped for by clinical medicine, generalizability simulated with random assignment in RCTs.

From Tacit Knowledge to Socio-Epistemic Bubbles in Science

Collins introduced a typology of tacit knowledge: *somatic*, *relational*, and *collective* (2010), which categorizes knowledge by where it resides. Exemplified with Polanyi's original example of "bike balancing," *somatic tacit knowledge* is encoded in bodies and brains, which

entails training to perform successful judgment and action. This encompasses cognitive skills such as chess-playing. *Relational tacit knowledge* is a matter of "how particular people relate to each other" (Collins 2010: 86). This may remain tacit due to intentional concealment, the practical costs of explication (Nelson 2003), or mismatched silence between insiders and outsiders (Collins 2010: 91-97). The last type, *collective tacit knowledge,* must be acquired through immersion in the collective practices and language of a community, as practical bike riding requires negotiating skills for right-of-passage with other vehicles and pedestrians from social cues, which makes it resistant to clear and decontextualized explication.

This distinction between *somatic* and the other two types where tacitness resides outside an individual is useful. But we argue for a new theoretical construct between *relational* and *collective.* Although Collins' typology may suffice for bicycle riding, it is ill-suited to capture the *tacit bubbles* underlying science in action. The acquisition and diffusion of tacit knowledge within *relational* or *collective* spaces are framed as binary—only you and your intimates know something or everyone in the field knows it.

The amateur epistemologist, General Donald Rumsfeld, who directed the Iraq War under U.S. President George W. Bush, articulated a related typology of knowns and unknowns involving the war. Like Collins, he too posited three categories: "known knowns" that the military could articulate and estimate, "known unknowns" that they could articulate but not estimate, and "unknown unknowns" about which they were entirely clueless. Scientific research often targets "unknowns," phenomena of great uncertainty, but the assessment of what is "known" among researchers is partial and often contested. As such, tacitness is not discrete but continuous: things may be more or less known.

We conceptualize science as a field of knowledge, eddied by *socio-epistemic bubbles* where implicit preferences, unarticulated beliefs, and unstated heuristics become shared among a local network of scientists. Tacitness in science resides above the individual but is neither ubiquitous across the collective nor confined within a clique of scientists. This continualizes the notion of an epistemic culture, which Knorr-Cetina articulated in the discrete contexts of molecular biology and particle physics (Knorr-Cetina 1999).

With respect to Rumsfeld's typology, note that his tripartite scheme missed a logical fourth category (Hann 2011): "unknown knowns" where knowledge is shared but unacknowledged and unrecognized. Within socio-epistemic bubbles, knowledge is obvious to those who populate that region of the scientific space, but obscure to those beyond it. Unlike a named discipline, the boundary of an epistemic bubble is typically invisible to both those on the in- and outside. What is more, these bubbles are continuous—each person experiences a different one as a function of their precise location and adjacencies.

This notion of an epistemic bubble is reminiscent of an "Invisible College," a dense community of "in-group" researchers who actively engage each other in studying a specialty (de Solla Price and Beaver 1966; de Solla Price 1961). Applying Collins' typology, a single epistemic bubble would correspond to the social space where *relational tacit knowledge* is shared and new collaborations emerge through homophily of research orientation (C. Zhang et al. 2018). Nevertheless, science comprises a continuous space that transcends a single community and contains innumerable, overlapping bubbles defined by those who have attended the same conferences, read the same articles, sat through the same seminars or lectures, used the same methods, and been party to the same conversations.

The space underlying *socio-epistemic bubbles* equates with the scientific field characterized by Bourdieu. This field reflects the distribution of scientific and cultural capital that shapes the *habitus* of scientists (Bourdieu 1991, 2004, 1975). We expand Bourdieu's discrete view of scientists' *habitus* to involve continuous gradation from scientist to scientist that spans tacit assumptions, heuristics, expectations, and speculations critical for the replication of scientific work. Bubbles in the scientific field are maintained through self-reinforcement by scientists who conduct, document, and share research with the scientific *habitus* that ultimately influence what can be precisely reproduced and replicated in fields of science.

Our analytical strategy turns this scientific field and associated socio-epistemic bubbles into a continuous metric space. The idea of deriving latent, abstract spaces from interdependent structured data for social research dates to Paul Lazarsfeld (Lazarsfeld and Henry 1968) and other network analysts who sought to capture latent representations of network data using algorithms like multidimensional scaling and block-models (Breiger, Boorman, and Arabie 1975). More contemporary efforts include work by Hoff and colleagues (2002), in which they propose a statistical method to generate low-dimensional features that constitute a "latent social space" wherein the relative position of individuals governs the formation of new ties. In all this work, the focus has been to construct low-dimensional embedding spaces that preserve distances between all networked entities with minimal distortion (Smith, Asta, and Calder 2019). In the decade, however, efforts to create relational spaces from networks have been revolutionized in terms of their scalability, resolution, and predictive power by the emergence of neural embedding models, which learn continuous vector representations of network data (Cui et al. 2019). Following this work, we use a neural embedding model to generate a manifold that succinctly encodes the complex relational structure underlying RCTs.

We use this approach, as detailed below, to represent the continuous latent space of socio-epistemic bubbles in which tacitness resides between the network of collaborating research scientists. Collaboration and co-authorship on a research project involve joint participation and can be derived from publication data following Breiger's approach to conceptualizing the social duality between people and groups (Breiger 1974). Substantively, a socio-epistemic bubble can be represented by proximity in the vector space of research co-authorship, which captures not only similarity in the choice of research topics and strategies (Foster, Rzhetsky, and Evans 2015), but also social obligations and cognitive preferences for certain approaches and answers (Teplitskiy et al. 2018). Recent research demonstrating that greater diversity in replication improves generalizability (Danchev, Rzhetsky, and Evans 2019; Belikov, Rzhetsky, and Evans 2022) suggests that social connections increase the number of unmentioned conditions and controls remaining tacit or implicit in print. This suggests that social proximity increases the likelihood of co-presence within a socio-epistemic bubble and the number of unarticulated invariances within experiments that defy straightforward replication.

The Rise of Evidence-Based Medicine

Pre-EBM period

The practice of RCTs sought to establish treatment efficacy and safety by eliminating potential confounding influences through random assignment. Popularized by Ronald A. Fisher during the first half of the 20th Century[5] (Fisher 1925, 1935), RCTs were adopted in medical science during the post World War II era to evaluate the efficacy of streptomycin to cure tuberculosis (Hill

[5] Prior to Fisher, Charles Pierce and his student Joseph Jastrow in 1884 reported their cognitive experiments of perceiving differences in weights using a randomization process (Peirce and Jastrow 1884). But Fisher's role in introducing and spreading the practice was essential (Hall 2007; Hacking 1988).

1952) and the polio vaccine (Meldrum 1998). In the 1950s, the German manufacturer Grünenthal introduced the animal-tested drug thalidomide for female morning sickness during pregnancy to the worldwide market. This was later linked to thousands of documented miscarriages and tens of thousands of severe congenital disorders worldwide. Even though a U.S. Food and Drug Administration (FDA) reviewer recommended against U.S. adoption, narrowly avoiding the tragedy, the event catalyzed U.S. regulation in the 1962 Kefauver-Harris amendments requiring drug manufacturers to obtain FDA approval based on multiple rounds of RCTs on drug efficacy and safety for human subjects before marketing them to U.S. physicians and the public (Junod 2008).

Despite the new regulation, diagnoses and interventions depended on individual physician's judgments, shaped through education and clinical experience. To address the continuing flow of heterogeneous medical studies and experience—including conflicting RCTs—with an institutional apparatus, the U.S. National Institutes of Health (NIH) organized the Consensus Development Conference Program in the 1970s. These conferences adapted the idea of a "Science Court" proposed by engineer Arthur Kantrowitz (Kantrowitz 1967) and set guidelines for Breast Cancer Screening in 1977, Drugs and Insomnia in 1983, Prevention and Treatment of Kidney Stones in 1988, and many more. Conferences convened panels of 10-20 experts on focused topics and engaged in deliberation (Solomon 2015). Despite the success of the consensus conference as a model and its diffusion around the world in the 1980s and 1990s (Solomon 2015), a wide variation of medical research and practices persisted in the U.S. and elsewhere (Wennberg 1984).

Evidence-Based Medicine and *Cochrane*

"Evidence-based medicine" (EBM) entered the medical lexicon in the early 1990s (Eddy 1990; Evidence-Based Medicine Working Group 1992). Its origination and spread can be attributed to the collective efforts of a core group of medical researchers, especially clinical epidemiologists in Canada and the U.K., who unleashed a scientific movement (Frickel and Gross 2005). Archibald Cochrane was a pioneering figure who championed the importance of accessibility to RCT results and the critical role of research synthesis in guiding medical practices (Cochrane 1972). His work inspired the founders of EBM, including Iain Chalmers, who led the Oxford Database of Perinatal Trials and the early Cochrane Collaboration, and David Sackett, who started the first EBM center at Oxford in 1994 and made methodological contributions with Gordon Guyatt and Brian Haynes at McMaster University in Canada (Au 2021).

Definitions of EBM emphasize the importance of incorporating "current best evidence" (Sackett et al. 1996) or "the best available scientific evidence" (Davidoff et al. 1995) in clinical decision making. Despite the rise of RCTs in medicine from the 1960s through the 1980s, the panelists of consensus conferences often favored their clinical experiences, delaying or negating the possibility of consensus (Daly 2005). The significant deviation of EBM from the previous approach, including the consensus conference model, was its prioritization of evidence obtained from RCTs. In other words, EBM prescribes a hierarchy of evidence, placing RCTs at the top, above observational studies, then conclusions from group deliberation with singular case studies at the bottom, subordinating judgment from experience and authority to numerical evidence generated from explicit experimental protocols. Under this evaluative regime, systematic reviews[6] are produced and clinical practice guidelines prepared to standardize medical interventions (Timmermans and Berg 2003).

[6] The term "systematic review" was first introduced in 1994 (Dickersin, Scherer, and Lefebvre 1994; Mulrow 1994), calling for combining traditional literature reviews with quantitative

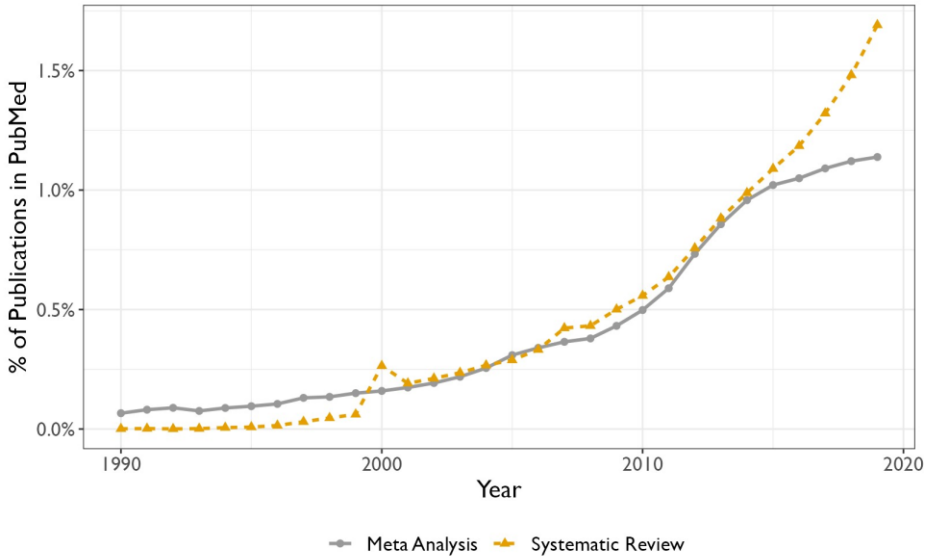
A crucial statistical tool for research synthesis used by EBM has been meta-analysis. Given the probabilistic nature of estimates from RCTs, a systematic review typically includes meta-analysis whenever multiple RCT-based studies are available to synthesize different estimates and produce overall conclusions harnessing increased statistical power. A biostatistician, William G. Cochran formalized the early statistical model in 1954 (Cochran 1954), but the term "meta-analysis" was coined by statistician Gene Glass (1976). The practice garnered attention in the 1980s and 1990s (Hunt 1997) and was adopted by biostatisticians in medical science (Shadish and Lecy 2015) and early advocates of EBM. The increased availability of electronic publication databases during the 1990s was essential in making meta-analysis feasible (Gurevitch et al. 2018; Timmermans and Berg 2003). Figure 1.1 displays the growing annual proportion of publications indexed as systematic reviews and meta-analyses in PubMed from 1990 to 2019.

The Cochrane Collaboration (now Cochrane) was founded in 1993 in London and named after Archibald Cochrane under the leadership of Iain Chalmers. Cochrane maintains 53 review groups across distinct medical areas such as pregnancy and childbirth, mental health, breast cancer, and many others, with more than 30,000 expert volunteers from health science worldwide (Cochrane, 2021). It has been the most prominent non-profit organization promoting EBM since its foundation (Salandra, Criscuolo, and Salter 2021; Jadad et al. 1998). The main channel of influence is their vast collection of systematic reviews on medical topics ranging from the efficacy of drugs to the effect of diagnosis, nutrients, exercise, and surgery to inform healthcare professionals, policymakers, and patients.

---

evidence to handle the fast-growing number of publications to inform evidence-based medical practice. A systematic review refers to a scholarly genre synthesizing literature on the diagnosis and treatments of specific clinical conditions (e.g., comparing the efficacy of chemotherapy and radiotherapy on breast cancer).

Figure 1.1: The Growth of Systematic Reviews and Meta-Analysis in PubMed



Socio-Epistemic Bubbles and RCTs

Notably, a persistent criticism against EBM contends that the application of clinical knowledge necessitates embodied judgment and experience residing in healthcare providers, highlighting the duplexity of medicine as "science" and "art" (Montgomery 2006). The diminishing authority of physicians and other healthcare providers has been a recent topic of debate, but tacit knowledge underlying the art and craft of medicine continues to persist through social institutions ranging from hospital wards to medical conferences (Menchik 2021).

Other critics of EBM have more directly raised questions about RCT's promise as a bias-free tool for evidence generation. At one extreme, the intrinsic impossibility of eradicating the effect of confounders due to RCT's probabilistic nature is pointed out (Worrall 2002). While this criticism touches on the epistemic nature of scientific experiments based on randomization, a case study of ten highly cited RCT-based studies reports that clinical trials involve a plethora of subtle and tacit decisions, including the initial sampling, gathering the baseline information, and appropriate level of randomization, quantifying the treatment outcomes (Krauss 2018). When RCTs are performed to evaluate non-pharmaceutical medical interventions (e.g., exercise,

nutrition, surgery), it is often challenging to maintain double-blindness as both clinicians and patients know what they do, which suggests that subconscious expectations of an intervention's therapeutic effect can play a role in conveying therapies and evaluating clinical outcomes (Worrall 2002). Moreover, it has also been reported that even the production of standardized clinical outcome measures and their interpretation in evaluating therapeutic effects can also be influenced by clinicians' education and experience (Greenhalgh et al. 2008). All these discussions all point to the existence of tacitness in RCTs.

We neither intend to argue that RCTs and meta-analysis using them are wrong (Ioannidis 2005) nor to document the complete list of latent sources causing variations in RCT or quantify their individual impact. We posit, however, that tacit aspects not explicated through publications but implicitly shared in the space of socio-epistemic bubbles can shape the results of RCTs. Specifically, we will demonstrate that decreased social diversity leads to increased homogeneity in effect direction and size. We do this by projecting RCTs collected in meta-analyses from Cochrane systematic reviews to the 'social space' inscribed by collaborating scientists. Insofar as RCT evidence must transcend its study environment to impact widespread medical practice and population health, tacitly shared research knowledge underlying homogeneous results should be accounted for in meta-analyses so that the health value of those findings is available to the public. We also demonstrate that later researchers outside the socio-epistemic bubble of early research on a medical practice are much more likely to reverse our understanding of that practice's medical value. These results suggest that tacit certainty may unintentionally decrease the RCT literature's relevance for population health.

Data and Setup

We construct our data from two sources for the following analyses: 1) RCT results collected in meta-analyses within Cochrane systematic reviews, 1997-2017, and 2) disambiguated author IDs and publication references identified in the PubMed Knowledge Graph (PKG). Using the PKG dataset, we construct a manifold vector space of the large biomedical literature and researchers using the collaboration of researchers manifested through co-authorship via authors and publication IDs.

Cochrane Systematic Reviews

Cochrane Database of Systematic Reviews (CDSR) is considered the most comprehensive and reliable systematic review for various healthcare settings and interventions (Salandra, Criscuolo, and Salter 2021; Jadad et al. 1998). CDSR was ranked 10th (among 165 journals) and 11th (among 167 journals) in the Medicine General and Internal category on the journal citation report from Clarivate in 2019 and 2020, respectively. Cochrane has also maintained an official partnership with the World Health Organization (WHO) since 2011, and most of the WHO guidelines cite systematic reviews (75% in 2015 and 90% in 2016) published by CDSR (Cochrane 2016).

The authors of Cochrane systematic reviews are unpaid experts in a given healthcare topic. They first develop a protocol for a systematic review (such as narrowing the scope of the topic and refining keywords and terms for search), then perform a comprehensive query through diverse electronic databases (e.g., PubMed, Embase) to find relevant studies for a given research topic and question. Once a pool of candidate studies is collected, review authors evaluate the relevance and validity of collected studies and decide which should be included (and excluded)

in the review. Reasons for exclusion range from high risk of bias or invalid study designs to unavailability of relevant clinical outcomes due to different study aims. Following this process, review authors occasionally conclude that no trustworthy RCT is available on a given topic. Review authors conduct meta-analyses when appropriate studies exist, usually with multiple clinical outcomes, and report results with systematic, quantitative reviews. This process allows us to test our question by associating the social proximity or density of RCTs on specific healthcare topics with homogeneity among their results.

Our subsequent analyses draw on 20,117 meta-analyses containing at least five studies we extracted from 1,962 Cochrane reviews published up to November 18, 2017. A single Cochrane Review typically covers multiple outcomes, such as different end time points, side-effects, etc., in order to evaluate the overarching effect of an intervention. Cochrane reviews can be updated when new eligible studies are found, but the identifiers of each review remain the same. Our data is drawn from the latest version as of November 18, 2017.

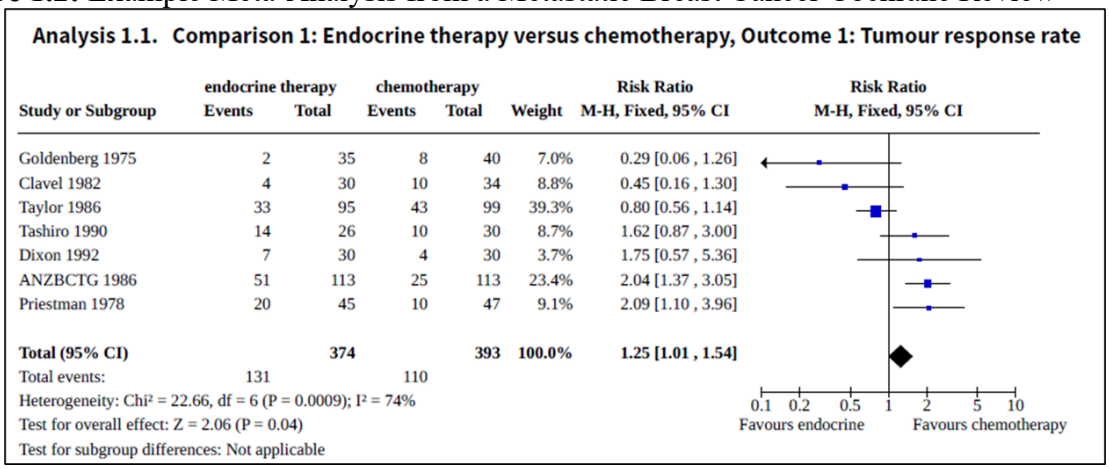
Reference sections from each review provide bibliographic information for studies listed under a "study heading," where review authors identify multiple publications within the same clinical trials or trial arms through a process of manual association. For example, a Cochrane review titled "Cranberries for preventing urinary tract infections" (Jepson, Williams, and Craig 2012)[7] identifies a clinical study, "Wing 2008," associated with two research articles published in 2008 and 2010. The 2008 article is titled "Comparison of urinary cytokines after ingestion of cranberry juice cocktail in pregnant subjects: a pilot study," published in the *Journal of Urology* by Drs. Deborah Wing, Pamela Rumney and colleagues from obstetrics and gynecology at the University of California, Irvine, School of Medicine[8] The 2010 article, "Daily cranberry juice for

[7] Cochrane Accession Number: CD001322 (Wilcken, Hornbuckle, and Ghersi 2003).
[8] PMID: 18707726.

the prevention of asymptomatic bacteriuria in pregnancy: a randomized, controlled pilot study," was also published in the *American Journal of Perinatology* by Drs. Wing, Rumney, and others.[9] The Cochrane review authors identified that the papers reported on the same underlying RCT and put them together. We harvested PMIDs associated with the associated studies and retrieved disambiguated author IDs and references for each article from the PKG dataset (Xu et al. 2020).[10]

Figure 1.2 displays an example meta-analysis from a Cochrane review, reporting the number of study participants, events, and estimates from seven trials comparing tumor response rates between endocrine therapy and chemotherapy on metastatic breast cancer.[11]

Figure 1.2: Example Meta-Analysis from a Metastatic Breast Cancer Cochrane Review



Following Cochrane conventions, we deem the subgroup meta-analyses for each clinical outcome as separate. Appendix Table A1.1 shows the number of systematic reviews and meta-analyses across the 52 Cochrane review groups we use for analysis. The breakdown of meta-

[9] PMID: 19562652.
[10] We use the second version of the PKG data for the disambiguated author IDs, covering MEDLINE indexed publications published by the end of 2019. We use the PMID-to-PMID citation table from the fourth version (C04_ReferenceList), which integrates PubMed's citation data, NIH's open citation collection, OpenCitations, and the Web of Science.
[11] Cochrane Accession Number: CD001322.

analyses by types of outcomes, measures, pooling methods, and model choices are documented in Appendix Table A1.2 (dichotomous outcomes) and A1.3 (continuous outcomes).

Encoding Socio-Epistemic Bubbles: Vector Representations of Publications in Social Space

Neural embedding models have become widely used to model and measure distance and change within language and networks. These models involve the construction of a dimensionalized vector space[12] in which geometrically proximate words or network nodes frequently share local linguistic or network contexts from training data. These were initially validated in the context of human language, where word proximities reflected underlying cultural meanings (Garg et al. 2018; Kozlowski, Taddy, and Evans 2019; Mikolov et al. 2013). This approach to continuous vector estimation from neural models has more recently emerged as a major approach for network analysis, dominating the performance of discrete network models and measures for prediction (Mikolov et al. 2013) .

Algorithmically, the accuracy and efficiency of Word2Vec's skip-gram architecture, which approximates the factorization of a text's word by context matrix (Levy and Goldberg 2014), increased its widespread popularity for modeling language and culture in the social sciences (Kozlowski, Taddy, and Evans 2019; Arseniev-Koehler and Foster 2020; Nelson 2021; Boutyline and Soter 2021; Lix et al. 2022). This approach generalized to networks by treating proximity between network nodes as those sharing network neighbors in the Deepwalk model

[12] These representations are considered "low-dimensional" from the perspective of the data on which they are estimated—there may be hundreds of thousands of unique words in a corpus or nodes in a network. Nevertheless, through cross-validation they often perform optimally with tens, hundreds or even thousands of dimensions, which is considered very "high-dimensional" from the perspective of social (and any formal) theory.

and its borrowed skip-gram architecture (Perozzi, Al-Rfou, and Skiena 2014), yielding separate embedding vectors for each node.[13]

Following Breiger's approach to conceptualizing the social duality between people and groups (Breiger 1974), we model the scientific field as a duality between scientist authors and the scientific artifacts (i.e., published papers) they produce. This approach to encoding the field of *socio-epistemic bubbles* enables us to capture similarities in research topics, preferences, and strategies among researchers and the distribution of scientific *habitus* (Foster, Rzhetsky, and Evans 2015) . Shared authorship indicates that co-authors assume credit and responsibility together, manifesting association and shared orientation. Moreover, the prevalence of broad co-authorship in medical research suggests collaborations are well-sampled. Nevertheless, this approach is limited in that it does not directly account for other social venues where socio-epistemic bubbles can emerge and reproduce, such as pedagogical contexts (Warwick 2011), laboratories (Latour and Woolgar 1979), or medical conferences (Menchik 2021).

We use the skip-gram approach that models such as Deepwalk also employs (Perozzi, Al-Rfou, and Skiena 2014), but we here adapt the Doc2Vec architecture (Le and Mikolov 2014) that can be used atop skip-gram to estimate vectors not only for authors but also for the papers they write together. In this way, Doc2Vec, with its recurrent neural network, enables us to place articles in terms of the authors who collaborated to write them. Specifically, we build a 100-dimensional social embedding space anchored by 8,359,189 disambiguated biomedical authors

[13] A subsequent node embedding approach, Node2vec (Grover and Leskovec 2016), propose alternative analytical approach that involves the generation of biased random walks across observed networks, flexibly allowing alternative parameterization to produce different metrics of similarity based on distinct hypotheses about the underlying topology of the network. Given the size and variability of our co-author network (~8M scientists across ~28M documents), we have no basis for specific hypotheses about network structure and so we use the assumption-free skip-gram/DeepWalk architecture.

with PKG, within which we assess the vector space position of 28,329,992 research indexed by PMID until the end of 2019. To include all the co-author and publication information in training, only authors who published more than one paper were considered.[14] We built our vector representations directly using the observed publication-author network without generating random walks based on it, considering the size of the network.[15]

We trained our vector space model with the Distributed Bag of Words (DBOW) approach, which uses the skip-gram architecture. The sliding window size that defines the size of training context for each author was set to 2,000—larger than the largest number of authors in a single published study. In this way, we ensure that in each training instance for each author and publication, the training context includes all co-authors, ignoring the sequential position of authors. This has the beneficial effect of linking first and last authors, who often work closely within a published project, no less closely than those arbitrarily beside one another in the author list. Consequently, the embedding learned by Doc2vec produces similar vector positions for authors who frequently co-author papers together. It also assigns similar vectors to articles co-authored by overlapping collaborators who share substantial tacit knowledge. For example, two postdocs or graduate students may never co-author a paper but connect indirectly through a shared principal investigator or through principal investigators who collaborate frequently with one another. By contrast, a large-scale RCT involving a unique collaboration among many otherwise disconnected authors would project to a sparse area of our embedding space between and distant from authors' prior work. This gap between the large RCT and author's prior studies would reflect a low likelihood of tacit knowledge transmission.

[14] The number of disambiguated author IDs from PKG 2020 v2 is 15,530,165 but removing the author IDs that appear only once drops the number of unique author IDs to 8,359,189.
[15] Perozzi et al. (2014) also suggest that applying the skip-gram model to non-random walks would be appropriate for web-scale large graphs.

We trained the model using 100 epochs, or iterations of training. In sum, this produces continuous, 100-dimensional representations for 28,329,992 PMIDs and 8,359,189 author IDs.[16] We validate the quality of the vector representations by attempting to retrieve learned article vectors from their composing author vectors across 20 random samples of 1,000 publications each. To do this, we take the author vectors associated with each publication, infer the position of a hypothetical publication they could have authored, and test its proximity to the original vector representation of the article written by those authors. Because it is intrinsically impossible to distinguish publications written by the same author(s), we evaluate the 1, 5, 10, and 20 most similar PMIDs from the inferred vector using cosine similarity. We find that it is possible to retrieve the target PMIDs with the rate of 65.26% (SD= 1.73), 86.16% (SD=1.06), 90.27% (SD=0.74), 92.9% (SD=0.77) within the pools of the top 1, 5, 10, 20 most similar documents, respectively. The sharp increase in self-retrieval for relaxed conditions implies that documents sharing authors are located close together in the 100-dimensional social embedding space. This confirms that continuous proximity or distance derived from this social space can reflect direct and indirect pathways of connection, communication, and latent expectation between the authors of biomedical research, increasing the likelihood that they reside in a socio-epistemic bubble of shared but unarticulated assumptions. Using this, we measure the proximity and density of RCTs curated in Cochrane reviews by mapping them onto the vector space of collaborating researchers and publications.

[16] We used the Python Gensim package (version 4.0) (Radim Rehurek 2010) to train the model.

Analyses and Results

We conduct four analyses to assess the influence of socio-epistemic bubbles underlying on published biomedical and healthcare RCTs. We first analyze the study-to-study pair. In this step, we examine the likelihood that estimates from two studies deviate from pooled estimates as a function of the proximity between the two studies measured across the social embedding space. In the second analysis, we examine the relationship between the overall heterogeneity of estimates at the meta-analysis level and density scores from author similarity across the studies. Thirdly, we perform a leave-one-out sensitivity analysis on all 20,117 meta-analyses to test whether the social density measure predicts conclusion invariance with the leave-one-out procedure. In our final analysis, we measure the distance between clusters of early versus later studies within the social embedding space and assess how social drift relates to contrasting conclusions. We present details of our analysis designs and results in the following subsections.

We use the "metafor" package in R (Viechtbauer 2010) to compute heterogeneity statistics and conduct the leave-one-out analysis, replicating the settings implemented in RevMan 5, the statistical software for meta-analysis supported by Cochrane (The Cochrane Collaboration 2020; Deeks and Higgins 2010).

Analysis 1: Social Proximity and Estimates Heterogeneity among Study Pairs

Our analysis starts with 1,279,974 pairs of estimates extracted from the 20,117 meta-analyses. The primary outcome is the level of heterogeneity among pairs of estimates. We employ Cochran's *Q* test statistic, which measures the degree of variability among individual estimates as the weighted sum of squared differences between individual estimates and the combined pooled estimates. Formally, $Q = \sum w_i(\hat{\theta}_i - \hat{\theta}_p)^2$, where $w_i, \hat{\theta}_i, \hat{\theta}_p$ are weights by

the pooling method, individual estimates, and the pooled estimate, respectively (Deeks and Higgins 2010).

The type of outcome measure—like risk ratio, odds ratio, or mean difference—and pooling methods (e.g. Mantel-Haenszel method or inverse-variance method) for a given meta-analysis are typically chosen by meta-analysts, the authors of Cochrane systematic reviews, reflecting characteristics of medical interventions and outcomes. We apply the identical set of outcome measures and pooling methods used for meta-analyses to compute Cochran's Q for each pair of estimates.

Under the null hypothesis, Q is assumed to be distributed following a chi-square statistic with $k - 1$ degrees of freedom, where k is the number of studies (thereby, the degree of freedom is always 1 for the pairwise analysis in this step). As Q manifests low statistical power with the small number of studies for detecting heterogeneity, a higher threshold than the conventional .05 is recommended for significance testing (West et al. 2011). Following the recommendation in the *Cochrane handbook for systematic reviews of interventions* (Higgins et al. 2019), we utilize a threshold of .10 ($p < .10$) to determine significant divergence in estimates between the two studies.[17] In other words, we denote that estimates from a pair of studies significantly deviate from the pooled estimate when the *p*-value of Cochran's Q is less than .10, which renders 184,007 of 1,279,974 pairs of estimates (14.38%) as highly heterogeneous following this procedure.

We measure the *social proximity* between the two clinical studies using the cosine similarity between the two study vectors from the social embedding space described above. When multiple publications are linked to a single clinical study, we obtain the centroid of all

[17] The .10 threshold is still a conservative choice considering k = 2 in this setting.

relevant publication vectors. The resulting study vector contains the mean value from all publication vectors along each dimension.

We consider two control variables in this analysis. We first aim to account for the potential impact of knowledge explicated through publication by assessing the degree of *reference overlap* between studies. This overlap is measured using the overlap or Szymkiewicz–Simpson coefficient, calculated as the number of references both studies cited (the size of intersection) divided by the smaller set of references from the two studies.[18] (Formally, $Overlap(X,Y) = \frac{|X \cap Y|}{Min(|X|,|Y|)|}$, where |X|=the size of set X.) When multiple publications are associated with a clinical trial, we assume the union of all referenced papers associated with all clinical trial publications indicates its prior knowledge stock.[19]

The second control variable addresses the temporal gap between every pair of studies, or *study year difference*, by taking the absolute difference between the years of study. A study year is typically identified from the study heading (for instance, Goldenberg 1975 in Figure 1.2 indicates 1975 as the study year), which is often the study's first or most substantial publication. In cases where a study heading lacks a specific year, we take the average of the publication years of associated research articles.

Table 1.1 presents descriptive statistics of the variables from 328,285 unique study pairs from 1,279,974 pairs, suggesting that the same study pair is compared an average of 4 times within Cochrane reviews on different clinical outcomes. Considering the hierarchical structure that pairs of estimates nest in meta-analyses, which recursively nest in systematic reviews, we

[18] Our results are similar if we use the Jaccard coefficient, defined as the intersection over the union of references between the two studies.
[19] When either of the references of the two studies could not be retrieved, we impute zero.

estimate the effect of social proximity on the heterogeneity using the multilevel logistic regression model that allows random intercepts for higher levels.[20]

Table 1.1: Descriptive Statistics of Variables from the 328,285 Unique Study Pairs

Variable Name	Mean	SD	1Q	Median	3Q	Min	Max*
Social Proximity	.554	.183	.454	.549	.658	-.360	1
Reference Overlap	.097	.150	0	.030	.146	0	1
Study Year Difference	7.065	6.631	2	5	10	0	80

Note: Social proximity and knowledge overlap become 1, which is the theoretical maximum when a systematic review makes a distinction between different trial arms reported in the same set of publications. The maximum value of 80 for the study year difference is observed in a systematic review that examined the effects of supplementation of vitamin A on maternal and newborn clinical outcomes (Cochrane accession number: CD008666), which includes a study published in *British Medical Journal* from 1931 (Green et al. 1931).

Eq. (1.1) describes the full model. It contains ν_{00k}, the review-level random intercepts and η_{0jk}, the random intercept for meta-analysis *j* nested in review *k*. Note that $H_{jik} = 1$ if a pair of estimates exhibits high heterogeneity (i.e., *p*-value of Cochran's Q is under .10), otherwise 0. The interclass coefficient from the fully unconditional model with three-level random intercepts is .421. This suggests considerable variability in effect heterogeneity across the meta-analyses and systematic reviews, supporting the use of the hierarchical model. In addition, the model incorporates contextual effects of *social density*, *reference density*, and *study year sparsity* with terms β_{01}, β_{02}, β_{03} by including the clustered means of each variable at the meta-analysis level. Before estimation, all variables are centered at their means and divided by the standard deviations reported in Table 1.1, which follows the logic of grand mean centering.

$$Logit(H_{ijk}) = \pi_{0jk} + \pi_1 Soc\_Proximity_i + \pi_2 Ref\_Overlap_i + \pi_3 Study\_Year\_Diff_i$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01} Soc\_Density_j + \beta_{02} Ref\_Density_j + \beta_{03} Study\_Year\_Sparsity_j + \eta_{0jk}$$

$$\beta_{00k} = \eta_{000} + \nu_{00k}$$

[20] We use the "lme4" package in R (Bates et al. 2007) for estimation.

$$Soc\_Density_j = \overline{Soc\_Proximity_{.jk}}$$
$$Ref\_Density_j = \overline{Ref\_Overlap_{.jk}}$$
$$Study\_Year\_Sparsity_j = \overline{Study\_Year\_Diff_{.jk}} \quad \cdots \text{Eq. (1.1).}$$

Table 1.2. Multilevel Logistic Model Estimates *from Analysis 1*

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Dependent Variable	Two Estimates Revealing Significant Heterogeneity				
Level-1 (Pairwise) Effects					
(Intercept)	-2.366*** (.029)	-2.357*** (.030)	-2.349*** (.029)	-2.351*** (.030)	-2.342*** (.030)
Social Proximity	-.045*** (.003)		-.044*** (.003)		-.037*** (.003)
Reference Overlap		-.043*** (.003)		-.041*** (.003)	-.034*** (.003)
Study Year Difference			.018*** (.003)	.010** (.003)	.009** (.003)
Contextual Effects					
Social Density	-.103*** (.026)		-.086** (.026)		-.086** (.027)
Reference Density		-.052* (.023)		-.018 (.026)	-.007 (.027)
Study Year Sparsity			.121*** (.031)	.130*** (.032)	.120*** (.032)
Group Random Intercepts	**Standard Deviation**				
Level-2: Meta-Analysis (Intercept)	1.198	1.198	1.198	1.199	1.198
Level-3: Review (Intercept)	.981	.983	.976	.977	.978
Deviance	898,504.5	898,504.4	898,453.5	898,476.5	898,341.2
Num. Obs.			1,279,974		
Num. Meta-Analysis			20,117		
Num. Reviews			1,962		

Note: All continuous variables are centered at their means and divided by the standard deviations reported in Table 1.1 for standardization. Accordingly, coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors are in parentheses.
* $p < .05$; ** $p < .01$; *** $p < .001$(two-tailed)

Table 1.2 shows the results from five models evaluating the effects of variables on the log odds of a study pair reporting highly heterogeneous estimates. Model 1 and Model 2 focus on *social proximity* and *reference overlap*; Model 3 and Model 4 add the impact of the *study year*

difference; lastly, Model 5 presents the estimates with all variables as described in Eq. (1.1). Note that contextual effects capture the difference between the clustered effects of variables at the higher and pairwise levels. The coefficients at level-1 estimate the impact of variables at the study pair level on the likelihood a study pair manifests a high level of heterogeneity, controlling for the *social density*, *reference density*, and *study year sparsity* of studies included in a meta-analysis.

Across all models listed in Table 1.2, the coefficients for level-1 variables suggest that closer social proximity or higher reference overlap decreases the odds that a pair of RCT-based estimates diverge. Unsurprisingly, *study year difference* increases the odds a pair of studies will significantly deviate from their pooled estimate. The positive correlation between *social proximity* and *reference overlap* (.327) and the negative correlation between *study year difference* and *social proximity* (-.194) and *reference overlap* (-.327) reduce the magnitude of estimates. Nonetheless, Model 5 demonstrates a significant impact of *social proximity*. Controlling for the covariates, the change of the *social proximity* from - 1 SD to + 1 SD translates[21] to a 7.1% decrease of the odds (=exp[-.037*2] - 1 ≈ -.071) that a pair of studies significantly disagree. Estimates from the fixed-effect model (Appendix Table A1.4) that controls for unobserved heterogeneity at the meta-analysis level are consistent, as shown in Table 1.2.

In Models 4 and 5, the contextual effect of *reference overlap* disappears, but other effects remain significant at the higher level. The coefficients blend higher-level effects of variables grouped at the meta-analysis (level 2) and systematic review (level 3). As the same study pair can appear multiple times across different meta-analyses within a review, we do not make a distinction between level-2 and level-3 contextual effects here. We note caution should be taken

[21] Using the mean and the standard deviation from Table 1.1, it would be from .371 (=.554 - .183) to .737 (.554 + 183), which is within the empirically possible range.

as the dispersion of clustered means is less than what Table 1.1 posts: standard deviations of *social density* and *study year sparsity* are .104 and 3, respectively, approximately half of the level-1 dispersion. Nevertheless, the negative coefficient of the *social density* indicates that a study pair is less likely to report substantially heterogeneous estimates if included in a meta-analysis containing clinical studies sourced from a socially dense research community. The positive coefficient of *time sparsity* tells a similar story: greater averaged time differences are associated with the likelihood of reporting more divergent RCT results.

Analysis 2: Social Density and Homogeneity within Meta-Analyses

We now zoom out from study pairs and investigate the association between increased *social density* and the overall heterogeneity of estimates collected in meta-analyses. To measure the overall dispersion of estimates, we employ the I^2 statistic (Higgins et al. 2019). Formally, the statistic is:

$$I^2 = Max[0, (\frac{Q - df}{Q}) * 100\%] \quad \cdots \text{Eq. (1.2).}$$

Q denotes the χ^2 statistics defined in Analysis 1, calculated as the sum of squared deviations of individual studies' estimated effect sizes from the pooled estimate; *df* is the degree of freedom or *k* - 1 where *k* is the number of meta-analysis studies (Higgins et al. 2003; Higgins and Thompson 2002). I^2 represents the "percentage of total variation across studies due to heterogeneity rather than chance" (Higgins et al. 2003; Higgins and Thompson 2002, 558). More intuitively, it captures the non-overlapping proportion of confidence intervals across estimates, independent of outcome measures type (e.g., odds ratios, risk ratios, mean differences) and number of studies included in the meta-analysis. The 25%, 50%, and 75% thresholds were initially proposed to mark low, moderate, and high levels of heterogeneity (Higgins et al. 2003),

but the 2019 second edition of the Cochrane Handbook revised the original thresholds, suggesting that: (1) from 0% to 40%, heterogeneity "might not be important"; (2) 30% to 60% "may represent moderate heterogeneity"; (3) 50% to 90% "may represent substantial heterogeneity"; and (4) 75% to 100% suggest "considerable heterogeneity" (Higgins et al. 2019). With this guide, we construct the dependent variable for the following analysis by categorizing the meta-analyses into low-level heterogeneity and at least moderate heterogeneity with 30% and 40% cutoffs. With the 30% and 40% cutoffs, 6,973 and 5,730 meta-analyses from the 20,117 (34.66% and 28.48%, respectively) represent at least moderate outcome heterogeneity. (See Figure A1.1 for the histogram of raw I^2 statistics across all 20,117 meta-analyses.)

The primary predictor for this analysis is *social density*, which we define as the average *social proximity* between unique study pairs within a meta-analysis as assessed in Analysis 1. This operationalization resembles a density measure for weighted undirected graphs, considering individual studies as nodes and social proximity as edge weights.

Similar to our previous analysis, we control for the impact of explicated knowledge by accounting for *reference density*, measured as the average *reference overlap* between study pairs per meta-analysis. This is comparable to our approach to measuring *social density.* We also account for time variation, assuming that greater temporal dispersion would lead to higher heterogeneity based on the result from Analysis 1. Unlike Analysis 1, in which we use the absolute differences of study years, here we employ the standard deviation of study years for each meta-analysis to control for the effect of *study year dispersion.* Additionally, we include the *number of studies* combined in a meta-analysis as a proxy for variability among study populations, although I^2 was designed to be independent of the number of studies pooled in a meta-analysis. We also consider the impact of *sample size dispersion* on I^2 by incorporating the

standard deviation of the number of subjects across studies. Lastly, we account for the impact of *types of outcome measure* (i.e., risk ratio, odds ratios, risk difference, mean difference, standardized mean difference) on the level of heterogeneity. Table 3 provides descriptive statistics for the covariates.

Table 1.3: Descriptive Statistics of the Variables from the 20,117 Meta-Analyses

Variable Name	Mean	SD	1Q	Median	3Q	Min	Max*
Social Density	.595	.125	.512	.582	.672	.099	.997
Reference Density	.154	.103	.077	.103	.167	0	.983
Study Year Dispersion	5.17	2.99	3.04	4.55	6.57	0	33.27
Number of Studies	9.09	7.33	5	7	10	5	141
Sample Size Dispersion	482.29	2,732.38	50.14	106.39	235.12	2.05	122,815.3

**Note*: The theoretical maximum for *social* and *reference density* is 1 when a systematic review makes a distinction between different trial arms reported in the same set of publications. The maximum value of 33.27 for *study year dispersion* is observed in a systematic review that examined the effects of supplementation of vitamin A on maternal and newborn clinical outcomes (Cochrane accession number: CD008666). The maximum value of 122815.3 for the *sample size dispersion* (i.e., standard deviation of the number of subjects) is from a review regarding the efficacy of the injected cholera vaccine (Cochrane accession number: CD000974), which includes multiple large-scale vaccination studies. The maximum value of 141 for the *number of studies* is from a Cochrane review studying the effects of placebos (Cochrane accession number: CD003974).

We continue to use the multilevel modeling approach to assess the impact of *social density* on meta-analysis heterogeneity. The model in this step comprises two levels: the first representing each meta-analysis and the second level the systematic review. Formally, the full model estimated is as follows:

$$Logit(L_{ij}) = \pi_{0j} + \pi_1 Soc\_Density_i + \pi_2 Ref\_Density_i + \pi_3 Study\_Year\_Disp_i + \pi_4 Num\_of\_Studies_i + \pi_5 Sample\_Size\_Disp_i + \pi_6 Odds\_Ratio_i + \pi_7 Risk\_Diff_i + \pi_8 Mean\_Diff_i + \pi_9 Standardized\_Mean\_Diff_i$$

$$\pi_{0j} = \beta_{00} + \beta_{01}\overline{Soc\_Density_{.J}} + \beta_{02}\overline{Ref\_Density_{.J}} + \beta_{03}\overline{Study\_Year\_Disp_{.J}} + \beta_{04}\overline{Num\_of\_Studies_{.J}} + + \beta_{05}\overline{Sample\_Size\_Disp_{.J}} + \eta_{0j}$$

L_{ij} = 1 if I^2 is lower than a threshold (i.e., $I^2 < .3$ or $I^2 < .4$), otherwise L_{ij} = 0 ···Eq. (1.3).

Eq. (1.3) allows variability among the estimates gathered in a meta-analysis to exhibit low heterogeneity with the systematic review level random intercept term η_{0j} for each review *j*. The interclass coefficients for the unconditional models with .3 and .4 thresholds are .316 and .336, respectively. The model includes contextual effects of the covariates at the systematic review level with β_{01}, β_{02}, β_{03}, β_{04}, β_{05} terms capturing the effects of clustered means for each covariate at the systematic review level. Consistent with Analysis 1, all covariates are centered at their means and scaled by their standard deviation as reported in Table 1.3 before estimation.

Table 1.4: Multilevel Logistic Model Estimates *from Analysis 2*

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Dependent Variable	Low Heterogeneity					
Threshold	L = 1 if $\boldsymbol{I^2}$ < 0.3, otherwise, L = 0			L = 1 if $\boldsymbol{I^2}$ < 0.4, otherwise, L = 0		
Level-1 Fixed Effects						
(Intercept)	.513*** (.048)	.510*** (.048)	.524*** (.048)	.937*** (.051)	.935*** (.051)	.944*** (.051)
Social Density	.154*** (.034)		.134*** (.035)	.177*** (.036)		.152*** (.037)
Reference Density		.095** (.032)	.038 (.034)		.123*** (.034)	.070 (.036)
Study Year Dispersion			-.075* (.031)			-.047 (.032)
Number of Studies	-.150*** (.020)	-.149*** (.020)	-.143*** (.020)	-.166*** (.021)	-.164*** (.021)	-.158*** (.021)
Sample Size Dispersion	-.071* (.034)	-.069* (.033)	-.073* (.038)	-.132** (.041)	-.127** (.040)	-.134** (.041)
Odds Ratio (OR)	.171* (.081)	.171* (.081)	.155 (.081)	.235** (.086)	.228** (.086)	.219* (.086)
Risk Diff. (RD)	.362* (.154)	.369* (.154)	.339* (.154)	.256 (.161)	.268 (.161)	.245 (.161)
Mean Diff. (MD)	.440*** (.062)	.431*** (.062)	.423*** (.063)	.274*** (.066)	.266*** (.066)	.263*** (.066)
Standardized Mean Diff. (SMD)	1.074*** (.092)	1.071*** (.092)	1.054*** (.092)	.975*** (.097)	.973*** (.097)	.965*** (.098)
Contextual Effects						
Social Density	-.075 (.053)		-.076 (.056)	-.107 (.056)		-.108 (.059)

Table 1.4 continued

Reference Density		-.067 (.052)	-.076 (.057)		-.060 (.055)	.052 (.061)
Study Year Dispersion			-.050 (.053)			-.036 (.056)
Number of Studies	-.072 (.086)	-.077 (.088)	-.083 (.088)	-.024 (.091)	-.012 (.093)	-.017 (.093)
Sample Size Dispersion	.051 (.047)	.054 (.047)	.059 (.048)	.106 (.056)	.106 (.055)	.113* (.056)
Group Random Intercepts	**Standard Deviation**					
Review Level	1.194	1.197	1.189	1.265	1.266	1.261
Deviance	22,967.7	22,982.7	22,950.4	21,142.9	21,154.9	21,131.2
Num. Obs.			20,117			
Num. Reviews			1,962			

*Note***:** All continuous variables are centered at means and divided by standard deviations in Table 1.3 for standardization. Accordingly, the coefficients indicate the change of log odds with the increase of unit standard deviation from Table 1.3. Standard errors are in parentheses.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table 1.4 presents estimates from six models. The first three utilize the .3 threshold, and the remaining the .4 threshold. Across all six models, coefficient π_4, representing the influence of the number of studies in a meta-analysis, is significantly negative. It suggests that as the number of studies in a meta-analysis increases, a higher level of heterogeneity in estimates is more likely, which we account for in this analysis. The impacts of *sample size dispersion* are consistently negative across all models, indicating that greater variation in study sample sizes within a meta-analysis is associated with a higher likelihood of estimates from meta-analyses being heterogeneous. The choice of outcome measures in each Cochrane review is made by the review authors, not the original study authors.[22]

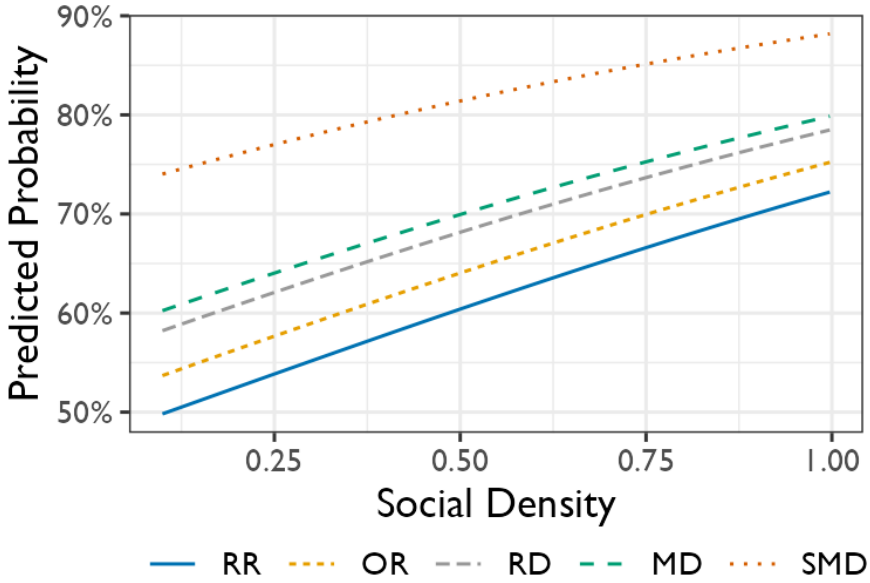
[22] The fixed effects model that employs review-level cluster-robust standard errors (presented in Table A1.5 in the Appendix) reports that only the standardized mean difference increases the likelihood of estimates being less heterogeneous. Thus, we do not provide a strong interpretation regarding the coefficients for the outcome measures.

The focal parameter in this analysis is π_1, which traces the impact of *social density* on heterogeneity. Models 1 and 4 display significant positive effects of *social density* on increasing the likelihood meta-analysis estimates are homogeneous. The coefficient for *reference density,* π_2, is positive and significant in Models 2 and 5, suggesting that higher *reference density* also predicts increased estimated homogeneity. *Reference density*'s correlation with *social density* is .369, and with *study year dispersion* -.384. Simultaneously including all three variables in models raises the specter of collinearity by inflating standard errors. Nevertheless, the impacts of *social density* hold consistent statistical significance across models.

Models 3 and 6 predict that a change of *social density* from -1 SD to +1 SD[23] translates to a 30.7% (=exp[.134*2] - 1 ≈ .307) or 35.5% (exp[.152*2] - 1 ≈ .355) increase in the odds of meta-analysis estimates exhibiting low heterogeneity, after accounting for the covariates. Appendix Table A1.5 reports estimates from the fixed-effect logistic model with specification from Eq. (1.3) but without contextual effects, arriving at similar results. To give a more concrete sense, Figure 1.3 visualizes the effect of *social density* in the probability terms based on Model 3, holding the values of other covariates at their global means shown in Table 1.3.

[23] With the mean and the standard deviation from Table 1.3, it would be from .470 (=.595 - .125) to .720 (=.595 + .125).

Figure 1.3: The Effect of Social Density on the Predicted Probability of the Estimates Combined in a Meta-Analysis Representing Low Heterogeneity ($\boldsymbol{I^2} < 0.3$)



Note: The solid lines show the marginal mean probabilities for a given meta-analysis to represent a low level of heterogeneity ($I^2 < 0.3$) based on Model 3 in Table 1.4, holding values for other covariates held at their global means reported in Table 1.3. RR, OR, RD, MD, and SMD refer to Risk Ratio, Odds Ratio, Risk Difference, Mean Difference, and Standardized Mean Difference, respectively.

Analysis 3: Social Density and Invariance using the Leave-One-Out Procedure

Sensitivity analysis for meta-analysis can take various forms. A meta-analyst may test and evaluate different models (e.g., fixed vs. random effects) using the same data. Furthermore, they may interrogate whether in/excluding studies disturbs the overall conclusion. The following applies a variant of the Leave-One-Out (LOO) analysis, akin to the approach initially proposed by Shenhav, Heller, and Benjamini (2015) as the meta-analysis *r-value*, with slight modifications we detail below. Suppose we have a meta-analysis comprising N studies, and the conclusion of the summary effect is statistically significant with the conventional 5% confidence level ($p < .05$). We then perform the same meta-analysis N times, iteratively excluding one study each time to maximize the *p*-value for the summary estimate with $N - 1$ studies. Conversely, if the *p*-value of the summary estimate with N studies is statistically insignificant ($p >= .05$), the same

iterative leave-one-out procedures are performed *N* times, aiming to minimize the *p*-value of the summary effect with $N - 1$ studies.

Note that Shenhav and colleagues (2015) primarily focus on the first *p*-value maximization but provide a generalized framework that allows more than a single study to be excluded (2015). While limiting the procedure to a leave-one-out scheme, we adapt their method. We operationalize that a meta-analysis manifests if (1) both the *p*-value of the original pooled estimate and the maximized *p*-value (comparable to the *r*-value) are below .05, or (2) both the original and the minimized *p*-values are equal to or above .05.

In contrast, we consider a meta-analysis does not reveal invariance if (3) the initial *p*-value is less than .05, but the maximized *p*-value is equal to or greater than.05, or (4) the original *p*-value is greater than or equal to .05, but the minimized *p*-value is less than .05. In other words, the procedure attempts to overturn the meta-analysis conclusion as much as possible through the removal of a single study.

Table 1.5 The Results of Applying the Leave-One-Out Procedure for 20,117 Meta-Analyses

	Summary Estimate $p >= 0.05$	Summary Estimate $p < 0.05$	**Total**
Invariant to LOO	9,005 (81.5%)	6,235 (68.2%)	15,240 (75.8%)
Variant to LOO	2,050 (18.5%)	2,827 (31.2%)	4,877 (24.2%)
Total	11,055 (100.0%)	9,062 (100.0%)	20,117 (100.0%)

Table 1.5 presents the results from this procedure applied to the 20,117 meta-analyses. We find that 18.5% of the meta-analyses with statistically insignificant summary estimates ($p < .05$) gain statistical significance, while 31.2% of those yielding statistically significant summary estimates ($p < .05$) becoming insignificant according to this procedure. This discrepancy is likely attributable to publication bias, wherein more statistically significant results

are published than insignificant ones. Regardless, the result suggests that the overall conclusion from a meta-analysis often hinges on in/exclusion of a single study.

An analogous hierarchical logistic regression framework is applied to assess the impact of social density on the invariance of meta-analysis to the LOO across 20,117 meta-analyses. We retain the same variables from Analysis 2 but introduce two additional controls: one indicator for model choice by review authors (i.e., random vs. fixed-effect model) and another to denote whether the *p*-value of the original summary estimate before the LOO is statistically significant ($p < .05$). The following equation, Eq. (1.4) describes our full model:

$$Logit(I_{ij}) = \pi_{0j} + \pi_1 Soc\_Density_i + \pi_2 Ref\_Density_i + \pi_3 Study\_Year\_Disp_i + \pi_4 Num\_of\_Studies_i + \pi_5 Sample\_Size\_Disp_i + \pi_6 Random\_Effect\_Model + \pi_7 p\_value\_under\_5\% + \pi_8 Odds\_Ratio_i + \pi_9 Risk\_Diff_i + \pi_{10} Mean\_Diff_i + \pi_{11} Standardized\_Mean\_Diff_i$$

$$\pi_{0j} = \beta_{00} + \beta_{01}\overline{Soc\_Density_{.j}} + \beta_{02}\overline{Ref\_Density_{.j}} + \beta_{03}\overline{Study\_Year\_Disp_{.j}} + \beta_{04}\overline{Num\_of\_Studies_{.j}} + \beta_{05}\overline{Sample\_Size\_Disp_{.j}} + \eta_{0j}$$

I_{ij} = 1 if a meta-analysis is robust to the LOO procedure, otherwise I_{ij} = 0 ··· Eq. (1.4).

Table 1.6 shows estimates from three models, with the last column reporting estimates for the full model defined in Eq. (1.4). Estimates for number of studies included in a meta-analysis (π_4) are positive across the models, reflecting the statistical property between sample size and results. The consistently negative coefficient for π_6 reflects different sensitivities toward the LOO procedure between significant and insignificant meta-analyses, shown in Table 1.5—significant summary estimates are more likely to be disturbed by the LOO procedure. Coefficient π_7 across all three models estimate negative, reflecting the random-effects model's production of larger standard errors than the fixed-effect model.

Table 1.6: Multilevel Logistic Model Estimates *from Analysis 3*

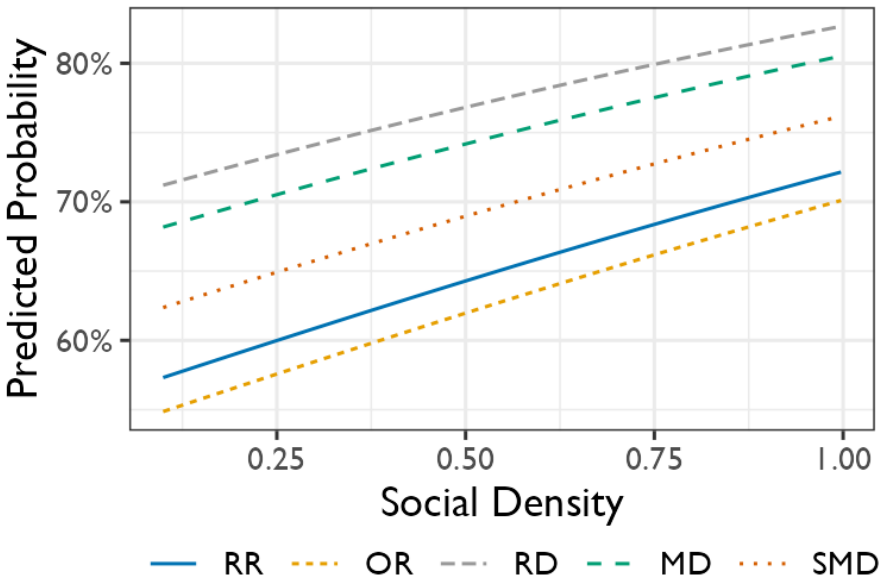
	Model 1	**Model 2**	**Model 3**
Dependent Variable	Invariant by the LOO Procedure		
Level-1 Fixed Effects			
(Intercept)	1.816***	1.806***	1.820***
	(.054)	(.054)	(.054)
Social Density	.087*		.091*
	(.035)		(.037)
Reference Density		.001	-.041
		(.033)	(.036)
Study Year Dispersion			-.059
			(.032)
Number of Studies	.376***	.368***	.376***
	(.031)	(.032)	(.032)
Sample Size Dispersion	-.060	-.058	-.061
	(.042)	(.042)	(.042)
Initial p-value < .05	-1.026***	-1.024***	-1.026***
	(.040)	(.040)	(.040)
Random Effect Model	-.139**	-.140**	-.137**
	(.051)	(.051)	(.051)
Odds Ratio (OR)	-.091	-.078	-.100
	(.073)	(.074)	(.074)
Risk Diff. (RD)	.628***	.619***	.610***
	(.164)	(.165)	(.165)
Mean Diff. (MD)	.476***	.469***	.467***
	(.062)	(.062)	(.062)
Standardized Mean Diff. (SMD)	.218**	.240**	.211*
	(.083)	(.083)	(.083)
Contextual Effects			
Social Density	.095*		.075
	(.049)		(.051)
Reference Density		.072	.038
		(.048)	(.052)
Study Year Dispersion			-.002
			(.049)
Number of Studies	.128	.128	.133
	(.072)	(.075)	(.074)
Sample Size Dispersion	.124*	.140*	.129*
	(.059)	(.060)	(.059)
Group Random Intercepts	**Standard Deviation**		
Review Level	.803	.821	.803
Deviance	20,731.4	20,761.8	20,724.6
Num. Obs.	20,117		
Num. Reviews	1,962		

*Note***:** All the variables are centered at the means and divided by the standard deviations in Table 1.3 for standardization. The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors are in parentheses.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

The impact of *social density* is again captured by π_1. Model 1 indicates its positive effect on invariance of conclusions to our LOO. Controlling for all the covariates, Model 3 shows the estimated coefficients of the full model specified in Eq. (1.4), revealing that increasing *social density* from -1 SD to +1 SD[24] can raise the odds of the pooled estimate being invariable to LOO by 20.0% (=exp[.091*2] - 1 ≈ .200). Fixed-effect model estimation using the cluster-robust standard error in Appendix Table A1.6 also demonstrates a similarly significant effect of *social density* on the invariance of meta-analysis pooled estimates.

Figure 1.4 displays the impact of *social density* on meta-analysis invariance to LOO in the probability terms based on Model 3, focusing on cases where meta-analyses were conducted by the random-effect model and pooled estimates are statistically significant ($p < .05$).

Figure 1.4: Effect of Social Density on the Predicted Probability of Meta-Analysis Summary Estimates Being Invariant to the Leave-One-Out procedure



Note: The focus is when the random-effects model is employed, and initial summary estimates are statistically significant ($p < .05$). The solid lines show the marginal mean probabilities for a given meta-analysis to be invariable with the leave-one-out procedure based on Model 3 from Table 1.6. Covariates are held at their global means, shown in Table 1.3. RR, OR, RD, MD, and SMD refer to Risk Ratio, Odds Ratio, Risk Difference, Mean Difference, and Standardized Mean Difference, respectively.

[24] The range would be from .470 to .720, the same as footnote 23.

Analysis 4: Shifts in Social Space and Diverging Evidence

In our final analysis, we adopt a temporal lens to investigate the link between shifts in the social positions of RCTs within our embedding space and the likelihood of changes in conclusions derived from the pooled estimates over time. To this end, we divide RCT studies from each meta-analysis into two temporal groups, applying their study years' ⅓ and ⅔ quantiles. For illustration, consider a hypothetical meta-analysis encompassing nine RCTs spanning each year from 1981 to 1989; we classify three studies from 1981 to 1983 as the 'early period,' and the last three from 1987 to 1989 as the 'later period.' Using this temporal slicing, we generate two summary estimates per meta-analysis—each exclusively pooling either 'early' or 'later' period RCTs.[25] We then assess whether the statistical conclusions from the later periods change or remain consistent compared to the early periods. (Seven meta-analyses were excluded for this analysis as they consist of RCTs with the same study year, rendering any temporal slicing impossible; hence, the subsequent analysis is based on 20,110 meta-analyses.)

We operationalize the summary estimates from later periods as temporally inconsistent with those from early ones when: (1) both summary estimates are significant but point to opposite directions—0.2% of the 20,110 meta-analyses; (2) a significant early period summary estimate ($p < .05$) turns to be insignificant ($p >= .05$) when only later periods studies are combined—14.7% of the 20,110 meta-analyses; or (3) an insignificant early period summary

[25] We calculate the ⅓ and ⅔ quantiles of study years per meta-analysis with linear extrapolation. RCTs with the same study years with a threshold are included when we compute summary estimates for each period. This ensures that 20,036 meta-analyses contain the same number of RCTs across early and later third periods. Nonetheless, 74 cases do not allow us to allocate the same number of RCTs across the early and later third periods. Consider a meta-analysis combining 10 RCTs, two studies from 1981, and eight studies from 1990: both ⅓ and ⅔ quantiles of study years are 1990. For those cases, we consider RCTs with the study year earlier than the ⅓ quantiles as the early period studies, otherwise later studies. As for the example, two RCTs with study years 1981 are considered early period studies and the other eight as later period ones.

estimate (p >= .05) gains statistical significance (p <.05) when only later periods studies are pooled—11.4% among the 20,110 meta-analyses.

In contrast, we categorize the cases as temporally consistent when: (4) both are statistically significant (p < .05), pointing in the same direction—17.3% of the 20,110 meta-analyses; or (5) both summary estimates from early and later periods are statistically insignificant (p >= .05)—53.4% of the 20,110 meta-analyses. In essence, a meta-analyst would reach the same conclusion over time for a consistent case such as (4) and (5), irrespective of whether they muster early or later samples of studies, while arriving at contrasting conclusions for inconsistent cases like (1), (2), and (3).

We aim to demonstrate how shifts in the position of RCTs in the social embedding space over time affect the consistency of statistical conclusions between early and later periods, which can yield conflicting healthcare recommendations. We obtain a centroid for the vector representation of RCTs for each period to characterize their positions in the social embedding space. Specifically, we compute the mean values for each dimension across all study vectors to locate centroids of early and later periods for each meta-analysis. We measure the cosine distance (1 - cosine similarity) between the early and later periods' centroids. We test whether this *centroid social distance* can predict the divergence of statistical evidence drawn from the two periods.

We utilize two-level logistic regression models again. In the full model, we consider controls similar to those from prior analyses—*reference overlap proportion*,[26] *mean study year*

[26] We compute this by dividing the number of unique papers referenced at least once by both the earlier and later studies by the total number of unique papers referenced by any of the earlier studies.

difference between the early and later periods,[27] *number of studies, mean sample size difference,*[28] *types of outcome measure*, and the model used for meta-analysis (i.e., *random effect* or *fixed effect model*). We also consider the statistical significance of summary estimates from the early period, termed *early significance*, in our modeling. We investigate how social proximity shapes whether scientific certainty shifts over time. As detailed in our theory section above, our hypothesis is that *social distance* between early and late investigators of the same medical practice will substantially increase the difference in their conclusions about the efficacy of that practice.

Specifically, we examine the effect of *centroid social distance* on statistical conclusions over time by interacting it with the variable *early significance*. If the interaction between social distance and early significance has a positive effect on temporal inconsistency, then greater social distance between early and later researchers will more likely reverse medical conclusions. If this is correct, then when early findings demonstrate decisive statistical significance about the benefits or drawbacks of a medical practice, later findings from distant researchers will more likely reverse early recommendations. Moreover, if the independent effect of social distance on temporal inconsistency also has a negative effect on temporal inconsistency, then social proximity between early and later researchers will more likely turn insignificant results into significant ones. If this is correct, when early findings about a medical practice are inconclusive, later results from socially proximate researchers will more likely confirm the early hunch, finding significant support for the practice. Table 1.7 presents the descriptive statistics for

[27] We obtain this by first taking averages of study years for early and later periods and then subtracting the earlier period's average study year from that of the later period.
[28] We also take the average number of participants in RCT studies from early and later periods and then obtain the absolute difference between them.

centroid social distance, reference overlap proportion, and *mean study year differences* between the early and the later periods from 20,110 meta-analyses.

Table 1.7: Descriptive Statistics of the Variables Based on the 20,110 Meta-Analyses

Variable Name	Mean	SD	1Q	Median	3Q	Min	Max*
Centroid Social Distance	.207	.112	.122	.191	.276	.002	1.195
Reference Overlap Proportion	.207	.149	.098	.183	.286	0	1
Mean Study Year Difference	9.50	5.81	5.40	8.33	12.30	.42	46.50
Mean Sample Size Difference	396.81	3,070.22	25.40	68.25	189.53	0	205,542

Note: The theoretical maximum for *centroid social distance* is 2. The maximum *sample size difference* is again observed from the systematic review of the efficacy of the injected cholera vaccine (CD000974), the same review in which the maximum *sample size dispersion is shown* in Table 1.3.

Eq. (1.5) describes the full model with the dependent variable indicating whether the statistical conclusions drawn from two periods differ (i.e., $I = 1$ for scenarios like (1), (2), and (3); otherwise, $I = 0$ for (4) and (5) detailed above).

$$Logit(I_{ij}) = \pi_{0j} + \pi_1 Centroid\_Soc\_Dist_i + \pi_2 Ref\_Overlap\_Prop_i + \pi_3 Mean\_Study\_Year\_Diff_i + \pi_4 Num\_of\_Studies_i + \pi_5 Mean\_Sample\_Size\_Diff_i + \pi_6 Random\_Effect\_Model + \pi_7 Odds\_Ratio_i + \pi_8 Risk\_Diff_i + \pi_9 Mean\_Diff_i + \pi_{10} Standardized\_Mean\_Diff_i + \pi_{11} Early\_Sig_i + \pi_{12} Centroid\_Soc\_Dist_i * Early\_Sig_i$$

$$\pi_{0j} = \beta_{00} + \beta_{01}\overline{Centroid\_Soc\_Dist_{.j}} + \beta_{02}\overline{Ref\_Overlap\_Prop_{.j}} + \beta_{03}\overline{Mean\_Study\_Year\_Diff_{.j}} + \beta_{04}\overline{Num\_of\_Studies_{.j}} + \beta_{05}\overline{Mean\_Sample\_Size\_Diff_{.j}} + \eta_{0j}$$

$I_{ij} = 1$ if a meta-analysis is robust to the LOO procedure, otherwise $I_{ij} = 0$ ··· Eq. (1.5).

Table 1.8 shows estimation results from three models using the variables described above. Model 1 demonstrated a significant positive effect of *centroid social distance* (π_1) on the likelihood of divergence in statistical conclusions between early and later summary estimates. The review-level contextual effect of centroid social distance, β_{01}, is also significant and positive. The substantial coefficient for *early significance* implies that initial healthcare

recommendations, supported by statistical evidence, are more likely to be reconsidered by researchers later examining the same or similar claims (Fanelli, Costas, and Ioannidis 2017).

Model 2 shows a significant crossover effect of *centroid social distance* over *early significance*. It unveils that earlier significant results are more likely to be challenged by distant groups of researchers later on, while earlier non-significant claims tend to gain statistical significance when later studied by researchers close to early study authors.

Table 1.8: Multilevel Logistic Model Estimates from *Analysis 4*

	Model 1	**Model 2**	**Model 3**
Dependent Variable	Inconsistent Conclusions between Early and Later Periods		
Level-1 Fixed Effects			
(Intercept)	-1.853***	-1.850***	-1.909***
	(.049)	(.050)	(.051)
Centroid Social Distance	.057*	-.111***	-.220***
	(.035)	(.031)	(.034)
Reference Overlap Proportion			-.001
			(.033)
Mean Study Year Difference			.096**
			(.034)
Number of Studies			-.265**
			(.030)
Mean Sample Size Difference			-.015
			(.028)
Random Effect Model	.303***	.304***	.314***
	(.049)	(.050)	(.050)
Odds Ratio (OR)	.094	-.099	.094
	(.071)	(.072)	(.072)
Risk Diff. (RD)	-.688***	-.700***	-.672***
	(.167)	(.169)	(.170)
Mean Diff. (MD)	-.372***	-.368***	-.375***
	(.060)	(.061)	(.061)
Standardized Mean Diff. (SMD)	-.241**	-.216**	-.198**
	(.082)	(.083)	(.083)
Early Significance	1.669***	1.692***	1.747***
	(.040)	(.040)	(.041)
Centroid Social Distance × Early Significance		.383**	.341***
		(.039)	(.050)
Contextual Effects			
Centroid Social Distance	.105*	.111*	.182***
	(.043)	(.044)	(.049)

Table 1.8 continued

Reference Overlap Proportion			-.049 (.050)
Study Year Difference			-.065 (.052)
Number of Studies			.003 (.073)
Sample Size Difference			.005 (.038)
Group Random Intercepts		**Standard Deviation**	
Review Level	.732	.745	.738
Deviance	20630.3	20530.2	20422.0
Num. Obs.		20,110	
Num. Reviews		1,962	

Note: All continuous variables are centered at the means and divided by the standard deviations. The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors are in parentheses. * p < .05; ** p < .01; *** p < .001 (two-tailed)

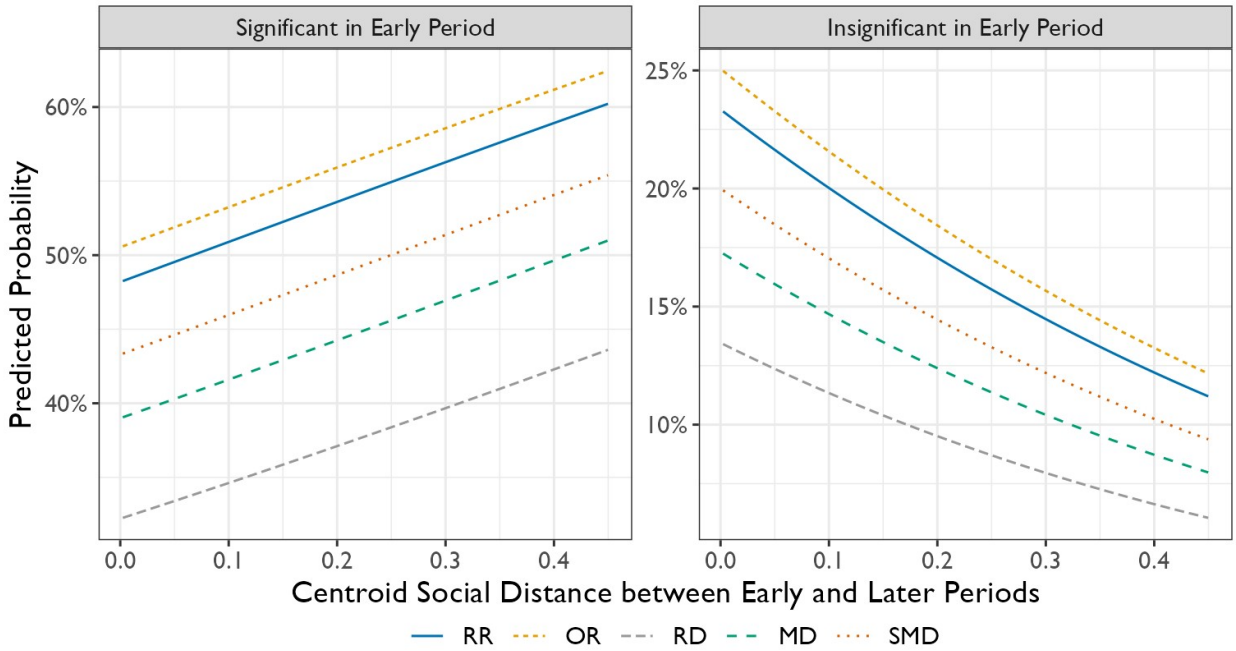
The last column of Table 1.8 presents the estimation results for Model 3, which estimates Eq. (1.5) incorporating all other control variables. Among controls, the *mean study year difference* (π_3) significantly influences the inconsistency between early and later period summary estimates. Unsurprisingly, a negative coefficient of the *number of studies* (π_4) implies that as a meta-analysis combines more studies, temporal inconsistency decreases. Importantly, the crossover effects of *centroid social distance* by *early significance* remain in the same direction as Model 2. Appendix Table A1.7 corroborates that employing review-level fixed effects yields similar results.

Results from Model 3 suggest that when the inference drawn from early studies is not significant ($p < .05$), a shift in *centroid social distance* from +1 SD to -1 SD[29] (i.e., increased social proximity between two periods) leads to a 55.3% ($\exp[-.220*-2] - 1 \approx .553$) increase in the odds of inconsistency between early and later period conclusions, with a statistically significant

[29] With the mean and the standard deviation from Table 1.7, it would be from .319 (=.207 + .112) to .095 (=.207 - .112).

later period estimate. Conversely, if the pooled estimate from early period studies is statistically significant ($p < .05$), an increase in *centroid social distance* from -1 SD to +1 SD raises the odds of inconsistency between the conclusions of early and later periods by 27.4% (exp[(-.220 + .341) * 2] - 1 ≈ .274) as summary estimates in the later period no longer maintain statistical significance. Based on Model 3, Figure 1.5 visualizes the varying effects of *centroid social distance* with *early significance* on the probability of inconsistency between evidence from the early and later periods, holding other continuous covariates at their global means and assuming a random-effects model is employed.

Figure 1.5: The Effect of Centroid Social Distance on the Predicted Probability of Inconsistency in Statistical Conclusions Between Early (First Third) and Later (Last Third) Periods



Note: The focus is when the random-effects model is employed. The solid lines show the marginal mean probabilities that a statistical conclusion from a later period manifests inconsistency to the early period for a given meta-analysis, based on Model 3 from Table 1.8. Covariates are held at their global means, shown in Table 1.7. RR, OR, RD, MD, and SMD refer to Risk Ratio, Odds Ratio, Risk Difference, Mean Difference, and Standardized Mean Difference.

The first scenario depicted in the left panel of Figure 1.5—where early significant findings are more likely to be challenged by more distant researchers—invokes two possibilities.

One modest possibility could be that the later-period researchers might inadvertently fail to satisfy certain tacit conditions required for the accurate replication of the initial claims. A more critical interpretation suggests that re-evaluation of the purportedly significant claims are more likely to be undertaken by researchers maintaining a degree of social independence from early-period researchers. Either interpretation confirms the expectation detailed above about how socio-epistemic bubbles might burst.

The second scenario from the right panel of Figure 1.5 illustrates how early non-significant claims tend to gain statistical significance when later examined by researchers close to early study authors. This reveals another implication of socio-epistemic bubbles. Socio-epistemic bubbles transmit tacit assumptions, beliefs, and expectations that influence subsequent examination of initial claims. Researchers within them likely share pragmatic goals and epistemic positions. Because researchers tend to utilize RCTs to report significant rather than insignificant results, reflecting widespread bias against the publication of null results (Rosenthal 1979; Pautasso 2010), later researchers close to the original studies eventually achieve evidence that early researchers sought unsuccessfully.

We select three cases that illustrate the role of social distance, each representing one of the three cases of temporal inconsistency in statistical conclusion as operationalized above as (1), (2), and (3). Rather than providing comprehensive validation with these cases, we seek to complement our quantitative analysis. The first illustrates the case in which both early and later period summary estimates manifest statistical significance but point in opposite directions. A Cochrane review entitled "Antibiotic prophylaxis versus no prophylaxis for preventing infection after cesarean section" (Smaill and Grivell 2014) evaluated the effects of various antibiotics.[30]

[30] Cochrane accession number: CD007482

One meta-analysis within the review focuses on the efficacy of Cefamycins in preventing maternal fever across nine studies. The conclusion drawn by combining early period studies (two studies from 1981 and another two from 1983) favors Cefamycins over no-antibiotics ($p < 0.001$). When only the later period studies (each from 1989, 1990, and 2001) are pooled, however, the recommendation flips, supporting no-treatment ($p < 0.01$). The centroid social distance between the two periods is .317 (the upper 83.9 percentile of cosine distance), illustrating the association between distance in social space and diverging conclusions between early and late studies.

The second illustrates how early significant results may be challenged by distant researchers as estimates from later periods studies do not find statistical significance. In a review titled "Antidepressants versus placebo for depression in primary care" (Arroll et al. 2009), the authors evaluate the efficacy of antidepressants versus placebo.[31] One analysis evaluates the efficacy of Tricyclic Antidepressants (TCAs) for depression symptoms 4 weeks after medication. Pooling three early period studies (1971, 1971, and 1979) would favor TCAs over placebo ($p < .01$) while three later period studies (1988, 1997, 1999) suggest no statistically meaningful differences between treatment and placebo groups (p=.119). Centroid social distances between early and late period studies is .310 (the upper 82.5 percentile of the centroid distance distribution).

Our final case illustrates a situation where early non-significant results gain statistical significance in later periods by researchers socially proximate to those who published early studies. In a review that assessed the side effects of Salmeterol in treating chronic asthma (Cates and Cates 2008), one meta-analysis estimates all drug-related adverse events.[32] Pooling five early

[31] Cochrane accession number: CD007954
[32] Cochrane accession number: CD006363

period studies (1992, 1994, 1998, 1998, 1998) leads to a statistically insignificant summary estimate ($p = .24$). Doing so with four later period studies (1999, 1999, 2000, 2004) yields a statistically significant summary estimate ($p < .05$), suggesting that Salmeterol may increase the likelihood of experiencing drug-related adverse events compared with the placebo group. The centroid social distance between the two periods is extremely low at 0.01, with 31% or 9 of the 29 authors involved in later studies overlapping those from the early ones. This suggests that later studies were performed by many of the same researchers to overcome the inclusiveness of early published findings.

Discussion

Summary of Results

Building on the theory of tacit knowledge, we develop the notion of a scientific field defined by overlapping socio-epistemic bubbles that capture the latent assumptions required for successful experimental replication. We then operationalize this by harnessing scientific publications encoded within neural embedding models to construct a continuous space within which bubbles of biomedical tacitness can be measured. Guided by the wisdom of crowds, our analysis shows that social proximity across clinical trials predicts an otherwise scientifically unaccountable consistency in meta-analysis estimates. The problem with this consistency is precisely its tacitness, bound up in shared assumptions and expectations, similarities of technique and interpretation that allow groups of socially proximate researchers to "find the same thing." This is a problem because shared tacit insight cannot transcend the particular RCTs performed and lift the efficacy of treatment in clinical practice. Instead, socio-epistemic bubbles may propagate false (un)certainty about medical efficacy and safety. Insofar as RCTs represent the pinnacle of

biomedical knowledge in the age of scientific medicine, unmeasured social proximity inscribing socio-epistemic bubbles wins the tribunals of science, but with less relevance to the clinic where lives hang in the balance.

In our analysis, we adapt the Doc2vec model to generate representational vectors for articles to capture the continuous space of socio-epistemic bubbles by encoding the collaborative pattern underlying biomedical research. We connect similarity-based measures from the social embedding space to estimates curated by Cochrane Reviews.

Our first analysis demonstrates that the closer pairs of clinical trials lie in social space, the lower the likelihood their estimates will deviate. In the second analysis, we investigate the relationship between social density, operationalized as the averaged cosine similarity between included studies, and the overall heterogeneity of meta-analysis estimates. We find that estimates collected from more dense social pools of publications tend to be markedly more homogeneous. In the third analysis, we apply a leave-one-out sensitivity analysis to assess the invariance of summary estimates. We test whether social density can predict the stability of statistical conclusions and find that summary estimates tend to be more invariable to sensitivity analysis when individual estimates are harvested from socially clustered publications. These results demonstrate the consistency of findings at the pairwise, aggregate, and marginal levels of analysis. Our final analysis offers a more nuanced picture of how socio-epistemic bubbles relate to shifting evidence from the research community. We demonstrate how early significant results are more likely to be later challenged by distant researchers than those close to early period researchers. Conversely, early null findings tend to gain support with statistical significance later when studied by researchers close to early studies authors.

In summary, we find that tacit knowledge consistently and substantially shapes certainty

regarding medical treatment from RCTs. This tacit confidence guides medical practices despite its inability to transcend the social-epistemic bubbles of research agreement that make it appear more efficacious and safef than can be communicated or impactfully diffused through clinical medicine for improved health.

Limitations

Our analyses have natural limitations. The first stems from the observational nature of our work. We repurposed existing data. While we have controlled for extensive covariates, our work admittedly remains associational. Nevertheless, we believe our results offer robust estimations across a range of modeling assumptions, including pairwise, aggregate, and marginal levels of analysis (i.e., analysis 1-3 above) and random vs. fixed-effect model implementations (see Appendix for Chapter 1).

We also note that our evaluation is drawn from clinical trials described in journal articles cataloged by MEDLINE. Our measurement strategy does not directly consider broader scientific contexts such as classrooms and conferences wherein tacit knowledge can flow and socio-epistemic bubbles grow. Moreover, our data did not cover results reported only in dissertations and unpublished work, which may manifest a higher proportion of failed studies (Goldacre 2014). In a similar vein, we excluded trials reported via publications from which we could not retrieve authors' identifiers from the PKG 2020 dataset. This is inevitable as the linkage between publications and systematic reviews depends on identifiers. Still, 89.83% (=30,660/34,133) of the unique RCT-review pairs from the 1,962 systematic reviews featuring 20,117 meta-analyses are linked with at least one PMID. Name-disambiguated author identifiers are available for most, but not all RCTs with PMIDs (96.07% of the 30,660 clinical trials). As publication bias may

nudge conclusions of published work toward greater significance, we suspect that a better linkage among clinical trials in social space would enhance the correlation between social proximity and estimates' homogeneity.

Another limitation stems from degrees of bias within clinical trials collected by Cochrane reviewers. For example, Cochrane reviews can include quasi-randomized clinical trials in which random allocation is conducted based on participants' date of birth or case record number. The authors of Cochrane reviews apply a guideline (i.e., GRADE: the Grades of Recommendation, Assessment, Development, and Evaluation) to distinguish clinical studies eligible for inclusion within a meta-analysis from those with high risk of bias slated for exclusion.[33] Our analysis assumes that assessments conducted by the Cochrane staff excluded clinical trials with a high risk of bias, but did not fully account for the proximity of studies in social space as we show here. That Cochrane Review authors already sought to cluster co-authored publications by identifying them as relevant to the same broader study, our analysis represents a conservative estimate of the effect of social proximity on published certainty.

Conclusion

RCTs have played a central role in contemporary medical science from the second half of the 20th Century in evaluating the efficacy and risks of medical treatment. Proponents of Evidence-Based Medicine legitimately embraced the idea of randomization to improve how medical interventions are assessed and practiced. The phrase "Evidence-Based" has gained authority not only in medicine but also in other social and policy-related sciences. Accordingly, the technique

[33] For instance, 1,962 reviews in our dataset initially identified 128,703 relevant studies, but 58.33% of the trials were not qualified to be included in the meta-analyses. The reasons for exclusion are not limited to a high risk of bias but span others, such as the unavailability of relevant clinical outcomes due to different study aims.

of meta-analysis has become a popular approach from which to draw generalizations from multiple experimental studies. Building beyond theories where tacit knowledge is assumed to universally permeate scientific subfields, we develop a theory of socio-epistemic bubbles through which tacit knowledge in science is shared with those connected and nearby, and not fully articulated within publications. We then interrogate whether the degree of social proximity between medical and healthcare RCTs can predict their dispersion of estimates. Our finding provides evidence that increased social proximity and density among epistemic bubbles of researchers is associated with under-dispersed estimates, which makes the overall conclusions drawn from meta-analysis appear more robust despite their limited ability to transcend the tacit configurations of RCTs.

We do not intend to argue that RCTs and meta-analysis are epistemologically weak or that they cannot contribute to better collective certainty regarding biomedical and other policy interventions. On the one hand, our analysis demonstrates the role of social embeddedness among researchers transferring tacit assumptions, techniques, and insights in the consensus-making process, even within otherwise controlled RCT settings. Conceptualizing scientific fields as continuous spaces of socio-epistemic bubbles, clinical trials may remain silent on experimentally relevant assumptions and protocols without intentional malpractice. On the other hand, our paper suggests that maintaining and enforcing a diverse research community would benefit biomedical science in several ways. First, it would prevent potential overconfidence produced by researchers densely linked within bubbles of shared intentions, assumptions, practices, and even expectations that drive social pressures for conformity. Second, if diverse and independent groups of researchers report comparable results, it would signal greater reliability and trustworthiness. Finally, it would provide opportunities for diverse biomedical researchers

and scientists to engage one another—neither generating unsustainable bubbles nor collapsing them—to clarify tacit aspects when significant heterogeneity is observed. This would motivate further research, the explication of scope conditions, and improved transfer to clinical practice and improved population health.

Our work makes distinctive contributions by bridging the social studies of scientific knowledge, metascience, and network science. Studies under the banner of "metascience" have uncovered critical, systematic sources like amplified effect sizes reported from early career researchers (Fanelli, Costas, and Ioannidis 2017) and scientists at U.S. institutions (Fanelli and Ioannidis 2013). While tacit knowledge has been articulated as playing an essential role in technology development (Collins 1974, 2010; Polanyi 1958), here we work out the sociological implication of its localization within socio-epistemic bubbles, especially regarding knowledge whose purported value lies in application outside the system in which tacit understanding is distributed and shared. We systematically demonstrate this phenomenon at scale in the context of healthcare RCTs. Our findings suggest that we may observe similar patterns in other fields ranging from fields of nonmedical biology to behavioral science, psychology, economics and related policy sciences that rely on randomized experiments to produce scientific and applied evidence.

Moreover, our analysis highlights the potential problem of under-dispersion as widespread and a new target for the field of metascience and the practice of statistical meta-analysis. These fields have historically focused on overdispersion and reducing upwardly biased confidence driven by low-powered studies. Our approach not only identifies the inflated certainty that may drive under-dispersion, but it enables the identification of socio-epistemological bubbles and a reasoned discounting of overconfidence by weighting studies

consistent not only with subject numbers, but potential social and epistemological proximity. We advocate for the use of network analytic concepts and methods within meta-analysis to account for latent social-epistemic structure within the realm of science. We hope that our approach can inspire further refinement and routine inclusion within meta-analyses to compensate for variation in agreement across RCTs.

Our work sheds light on an important but neglected reason why performing experiments and meta-analyses that aggregate them cannot mechanically resolve and harmonize scientific disagreement and dispute, contrary to popular expectations. This resonates with Karl Mannheim's proposed "task of solving the problem of the social conditioning of knowledge by boldly recognizing these relations and drawing them into the horizon of science itself and using them as checks on the conclusions of our research" (Mannheim [1936] 1991, 237). By not ignoring latent social bubbles of self-reinforcing agreement, but measuring them, we illustrate how biomedical knowledge can be better calibrated and improve its translation into health. Furthermore, explicit measurement of the social landscape underlying current biomedical understanding can allow us to design the diversity required to improve it.

Appendix For Chapter 1

Table A1.1: Distribution of the Number of Systematic Reviews across Disease Categories

Category	# Reviews	# Meta-analysis
Pregnancy and Childbirth	156	1,909
Airways	109	1,058
Neonatal	80	680
Heart	78	1,300
Common Mental Disorders	76	1,085
Kidney and Transplant	75	858
Pain, Palliative and Supportive Care	74	640
Gynaecology and Fertility	73	515
Stroke	72	559
Hepato-Biliary	70	1,403
Schizophrenia	63	646
Anaesthesia	59	727
Musculoskeletal	59	505
Colorectal	55	415
Acute Respiratory Infections	50	462
Infectious Diseases	47	428
Metabolic and Endocrine Disorders	45	690
Emergency and Critical Care	38	394
Gynaecological, Neuro-oncology and Orphan Cancer	38	408
Upper GI and Pancreatic Diseases	38	346
Vascular	37	293
Injuries	36	420
Tobacco Addiction	35	306
Wounds	32	99
Bone, Joint and Muscle Trauma	31	288
Drugs and Alcohol	31	328
Eyes and Vision	29	116
Hypertension	28	408
Oral Health	28	96
Breast Cancer	27	229
Developmental, Psychosocial and Learning Problems	25	399
Dementia and Cognitive Improvement	24	110

Table A1.1 continued

Effective Practice and Organisation of Care	22	121
Incontinence	21	103
Inflammatory Bowel Disease	21	146
Back and Neck	19	191
Consumers and Communication	17	191
ENT	16	86
Haematology	15	342
Skin	15	65
Epilepsy	14	170
Urology	12	66
HIV	10	71
Movement Disorders	10	121
Fertility Regulation	9	53
Multiple Sclerosis and Rare Diseases of the CNS	8	33
Neuromuscular	8	49
Sexually Transmitted Infections	7	19
Lung Cancer	6	111
Work	6	12
Cystic Fibrosis and Genetic Disorders	4	14
Public Health	4	33
Total	**1,962**	**20,117**

Note: The categories were classified according to the Cochrane Review Groups that specialize in each disease category. The names of the groups have been evolving. The most up-to-date group name in April 2021 is used in Table A1.1. For example, the "Back and Neck" group was first registered as the "Back Review Group for Spinal Disorders" in Dec 1998, and the group changed its name as "Cochrane Back Review Group" in 1999 and started to use the current name, "Cochrane Back and Neck", in 2015.

Table A1.2: Breakdowns of Dichotomous Outcomes

Outcome Measure	Pooling Method	Model		Total
		Fixed-Effect	Random-Effect	
Odds Ratio	M-H	1,016	950	1,966
	Inverse Variance	9	41	50
Peto Odds Ratio	PETO	973	-	973
Risk Ratio	M-H	5,050	5,169	10,219
	Inverse Variance	117	327	444
Risk Difference	M-H	170	296	466
	Inverse Variance	-	2	2
Total		**7,335**	**6,785**	**14,180**

Table A1.3: Breakdowns of Continuous Outcomes.

Outcome Measure	Model		Total
	Fixed-Effect	Random-Effect	
Mean Difference	1,312	2,652	3,964
Standardized Mean Difference	489	1,484	1,973
Total	**1,801**	**4,136**	**5,937**

Table A1.4: Fixed Effects Model Estimates of Eq. (1.1), with Level-1 Effects

	Model 1	Model 2	Model 3	Model 4	Model 5
Social Proximity	-.046***		-.045***		-.037***
	(.009)		(.009)		(.010)
Reference Overlap		-.044***		-.041***	-.035***
		(.008)		(.008)	(.009)
Study Year Difference			.018*	.010	.009
			(.009)	(.009)	(.009)
Num. Obs.			1,279,974		
Num. Meta-Analysis			20,117		
Num. Reviews			1,962		

Note: The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors inside the parentheses indicate cluster-robust standard errors at the meta-analysis and the unique study pairs in a review.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed t-tests).

Table A1.5: Fixed Effects Model Estimates of Eq. (1.3)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Threshold	L = 1 if $I^2 < 0.3$, otherwise, L = 0			L = 1 if $I^2 < 0.4$, otherwise, L = 0		
Social Density	.160**		.138*	.184***		.157**
	(.052)		(.055)	(.053)		(.056)
Reference Density		.101*	.042		.130*	.074
		(.048)	(.054)		(.053)	(.058)
Study Year Dispersion			-.082			-.053
			(.054)			(.055)
Num. of Studies	-.158**	-157**	-.150**	-.174***	-.171***	-.165***
	(.055)	(.055)	(.055)	(.050)	(.050)	(.050)
Sample Size Dispersion	-.082	-.079	-.084	-.152*	-.147*	-.154*
	(.059)	(.057)	(.059)	(.077)	(.075)	(.076)
Odds Ratio	.146	.135	.139	.224	.206	.216
	(.264)	(.264)	(.265)	(.272)	(.273)	(.274)
Risk Diff.	.324	.336	.321	.166	.181	.166
	(.414)	(.414)	(.413)	(.399)	(.398)	(.398)
Mean Diff.	.265	.255	.257	.090	.078	.083
	(.154)	(.152)	(.153)	(.154)	(.152)	(.153)
Standardized Mean Diff.	1.068***	1.054***	1.059***	.943**	.925**	.935**
	(.310)	(.309)	(.307)	(.306)	(.306)	(.305)
Num. Obs.			20,117			
Num. Reviews			1,962			

Note: The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors inside the parentheses indicate cluster-robust standard errors at the systematic reviews. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed t-tests)

Table A1.6: Fixed-Effects Model Estimates of Eq. (1.4)

	Model 1	**Model 2**	**Model 3**
Dependent Variable	Invariant by the LOO Procedure		
Social Density	.093* (.046)		.098* (.47)
Reference Density		.003 (.044)	-.046 (.045)
Study Year Dispersion			-.065 (.044)
Num. of Studies	.406* (.163)	.398* (.163)	.407* (.165)
Sample Size Dispersion	-.072 (.068)	-.069 (.067)	-.072 (.068)
Initial *P*-value < .05	-1.191*** (.082)	-1.191*** (.082)	-1.191*** (.082)
Random Effect Model	-.193* (.093)	-.195* (.093)	-.190* (.092)
Odds Ratio	-.302 (.195)	-.306 (.195)	-.304 (.195)
Risk Diff.	.598* (.302)	.605* (.302)	.592 (.303)
Mean Diff.	.393*** (.118)	.387** (.118)	.390*** (.118)
Standardized Mean Diff.	.552** (.182)	.550** (.181)	.552** (.181)
Num. Obs.	20,117		
Num. Reviews	1,962		

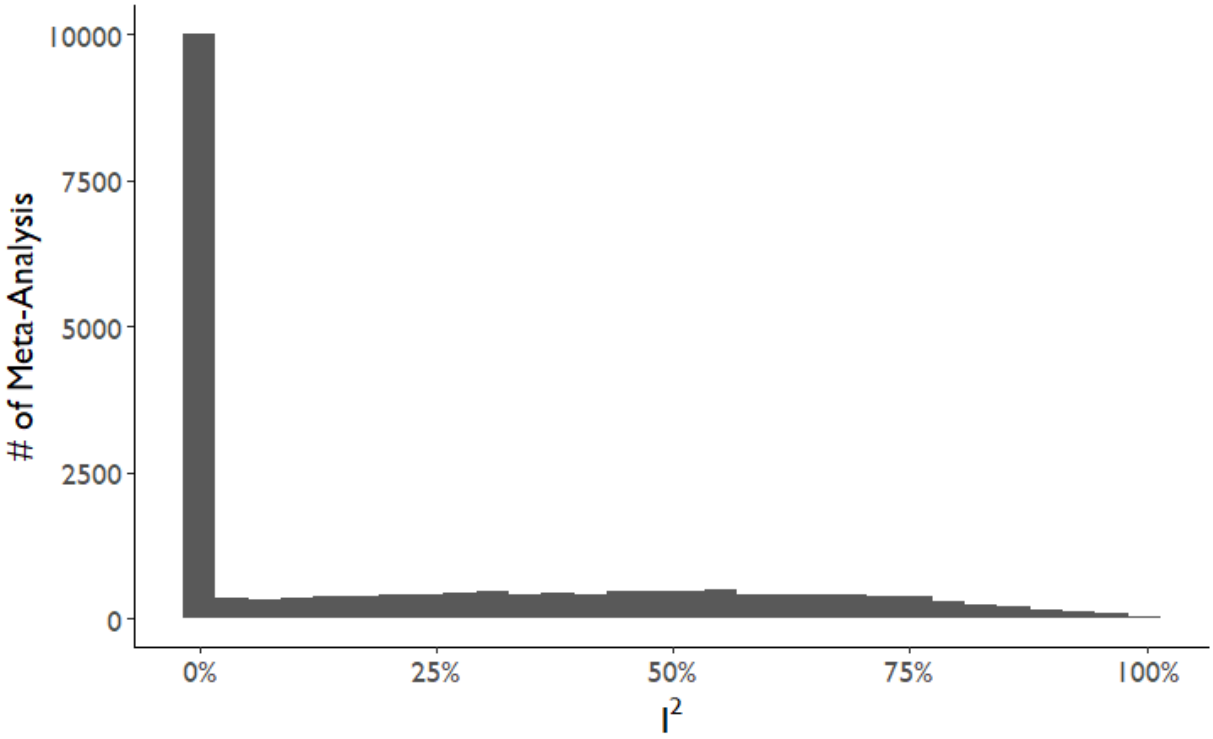
Note: The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable reported in Table 1.3. Standard errors inside the parentheses indicate cluster-robust standard errors at the systematic reviews. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed t-tests).

Table A1.7: Fixed-Effects Model Estimates of Eq. (1.5)

	Model 1	**Model 2**	**Model 3**
Dependent Variable	Inconsistent Conclusions between Early and Later Periods		
Centroid Social Distance	.062 (.040)	-.142** (.055)	-.260*** (.059)
Reference Overlap Proportion			-.004 (.039)
Mean Study Year Difference			.102* (.045)
Number of Studies			-.277** (.085)
Mean Sample Size Difference			-.021 (.047)
Random Effect Model	.286** (.101)	.294** (.104)	.303** (.104)
Odds Ratio (OR)	.232 (.180)	.241 (.177)	.225 (.176)
Risk Diff. (RD)	-.733* (.321)	-.728* (.299)	-.718* (.302)
Mean Diff. (MD)	-.264* (.128)	-.256* (.129)	-.275* (.119)
Standardized Mean Diff. (SMD)	-.333 (.197)	-.297 (.195)	-.272 (.188)
Early Significant	1.763*** (.118)	1.815*** (.118)	1.870*** (.111)
Centroid Social Distance × Early Significant		.474*** (.089)	.431*** (.090)
Num. Obs.		20,110	
Num. Reviews		1,962	

Note: The coefficients indicate the change of log odds with the increase of unit standard deviation used to standardize each variable. Standard errors inside the parentheses indicate cluster-robust standard errors at the systematic reviews. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed t-tests).

Figure A1.1: Histogram of I^2 Statistics from 20,117 Meta-Analyses



Chapter 2

Limited Diffusion of Scientific Knowledge Forecasts Collapse*

Abstract

Market bubbles emerge when asset prices are driven unsustainably higher than asset values and shifts in belief burst them. We demonstrate the same phenomenon for biomedical knowledge when promising research receives inflated attention. We predict deflationary events by developing a diffusion index that captures whether research areas have been amplified within social and scientific bubbles or have diffused and become evaluated more broadly. We illustrate our diffusion approach contrasting the trajectories of cardiac stem cell research and cancer immunotherapy. We then trace the diffusion of unique 28,504 subfields in biomedicine comprising nearly 1.9M papers and more than 80M citations and demonstrate that limited diffusion of biomedical knowledge anticipates abrupt decreases in popularity. Our analysis emphasizes that restricted diffusion, implying a socio-epistemic bubble, leads to dramatic collapses in relevance and attention accorded to scientific knowledge.

* Co-authored with James A. Evans, Department of Sociology, University of Chicago and Santa Fe Institute; and Robert S. Danziger and Jalees Rehman, Department of Medicine, University of Illinois at Chicago. Forthcoming in *Nature Human Behaviour*. Text and figures are reused with permission from Springer Nature.

Introduction

Market bubbles emerge when widespread opinions about an asset, such as housing or securities, create self-reinforcing information that drives its price much higher than its value to society (Arthur 1995). These bubbles are characterized by a swift surge in popularity, fueled by beliefs that the value may continue to rise and persist, leading to speculation. Such bubbles burst when shifts in opinion, often catalyzed by new data or events, precipitate radical discounts in pricing (Harras and Sornette 2011). Science observers and researchers themselves have drawn parallels in science (Goldman and Shaked 1991; Pedersen and Hendricks 2014; Evans et al. 2011), which involves considerable investment in capital, attention, and other resources based on highly uncertain knowledge about the outcomes of research. This exposes science to the risk of forming bubbles analogous to financial markets (Pedersen and Hendricks 2014; Evans et al. 2011). Here, we operationalize the concept of scientific bubbles and their collapse, proposing a measurement framework and demonstrating that ideas and findings in science can experience abrupt booms and busts of popularity and credibility that may yield adverse consequences for science and scientists alike.

In the system of biomedical knowledge, citation counts have come to function as an operational currency (Fortunato et al. 2018; Partha and David 1994), serving as a measure of the importance and impact of scientific work. This is also reflected by increasing interest in the development of indicators tracing emergent, disruptive, or breakthrough science and technology (Small, Boyack, and Klavans 2014; Funk and Owen-Smith 2016; Klavans, Boyack, and Murdick 2020; Weis and Jacobson 2021; Lin, Evans, and Wu 2022), which typically incorporate citation counts as key components. The citation metric manifests some distortion, however, from the inflation of citation counts with historical growth in articles (Petersen et al. 2019) and the

unequal size of fields (Hutchins et al. 2016). Inspired by the analogy between financial and scientific bubbles, here we forecast substantial and dramatic declines in the popularity of research ideas—the bursting epistemic bubbles—as the degree to which those ideas remain concentrated within the same collection of authors, institutions, and biomedical subfields, failing to diffuse across social and scientific space despite initial popularity. We argue that this limited diffusion may indicate inflated attention to particular ideas that may not generalize or withstand broader scrutiny, ultimately leading to disappointment and disillusionment within the scientific community.

Consider the extreme but illuminating case of cardiac regeneration in biomedicine. Dr. Piero Anversa and collaborators led research in cardiac regeneration at the turn of the 21st Century by asserting the possibility of damaged heart muscle tissue after myocardial infarction with stem cells and progenitor cells drawn from the bone marrow or within the heart (Taylor and Heath 2022). During Anversa and collaborators' peak productivity, they also exercised significant influence over the research narrative, sitting on editorial boards of high-profile American Heart Association journals like *Circulation Research* (Dr. Anversa alone reviewed hundreds of papers for *Circulation Research*, more than any other researcher in this period), serving on the NIH National Institute on Aging's Board of Scientific Counselors (2008-2013) and an interlocking matrix of NIH grant review panels. Nevertheless, findings from early cardiac regeneration work not only failed to generalize, but the experiments could not be replicated by other researchers (Murry et al. 2004). This resulted in a dramatic breach of trust, the retraction of more than 30 related papers from leading journals, a marked discount in citations to the subfield, diminished confidence in the near-term prospects of cardiac regeneration, and Anversa's forced departure from Harvard. This, in turn, adversely impacted even those researchers who had been

studying cardiac regeneration using more rigorous scientific approaches who had identified reproducible mechanisms underlying the phenomenon (Osafune et al. 2008).

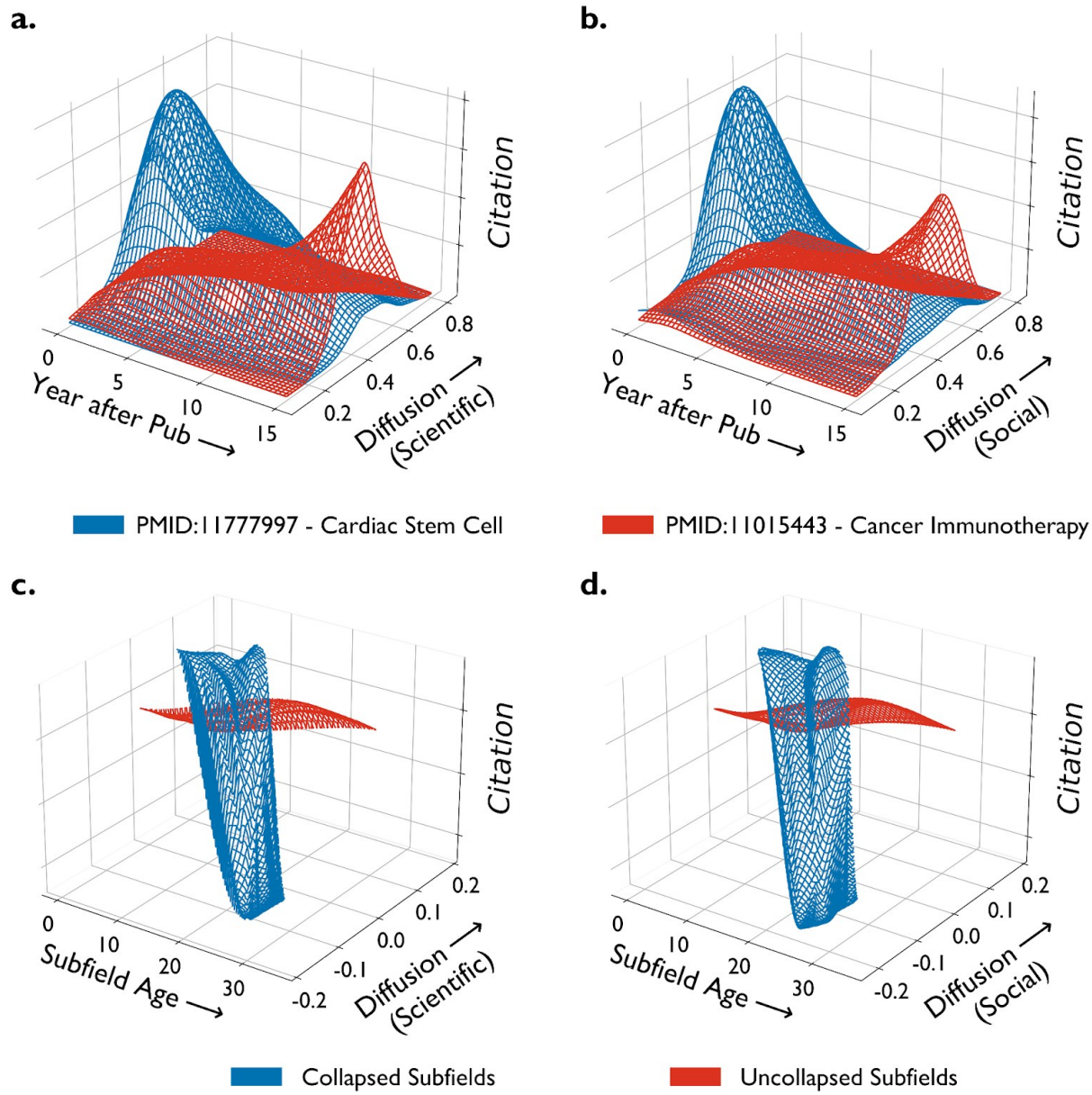
Our approach, however, aims to generalize beyond the severe research misconduct of an individual or a team of scientists. Accurate and honestly reported medical findings can still fail to generalize beyond the specific context of their initial investigation, despite optimism and hype regarding their transformative potential for medicine. More critically, as highlighted by science commentators (Harris 2017) and biomedical researchers (Neimark 2015; Hughes et al. 2007), unintended collective failures can also occur, as exemplified by the widespread use of misidentified or contaminated cell lines contributing to unjustified hype and misdirected attention and resources in the field. This phenomenon suggests the need for a more refined and multi-faceted framework to better model and evaluate the trajectories of scientific attention.

In this study, we demonstrate that fragile and overhyped biomedical findings could have been anticipated by analyzing their diffusion through the system of science. Utilizing PubMed Knowledge Graph (Xu et al. 2020), a large-scale bibliographical database, we provide a framework that considers distances between publications and their citing papers within the "scientific space" constituted by co-investigated biomedical entities and the "social space" constituted by collaborating scientists. Specifically, we develop a diffusion index to capture whether ideas have been amplified within social and scientific bubbles (Teplitskiy et al. 2018), or diffused more widely and tested for robustness across diverse research communities (Belikov, Rzhetsky, and Evans 2022). This approach allows us to gain insight into the diffusion of research ideas and their impact, ultimately helping us more rapidly assess the value and potential of scientific findings.

Our work demonstrates how a lack of diffusion measured by this framework—indicative of the existence of a scientific bubble—can anticipate a rapid decline in popularity as confidence bubbles burst. Applying the conceptual and measurement tools detailed below (Methods), we first compare two distinct trajectories from cardiac stem cell and cancer immunotherapy research papers. The upper panels of Figure 2.1 illustrate the approach with two contrasting papers. Figure 2.1.a and 2.1.b depict the diffusion and citation trajectories of an early paper (Quaini et al. 2002) from Dr. Anversa's group on cardiac muscle regeneration using bone-marrow-derived cells and a seminal paper on cancer immunotherapy conducted by Dr. Honzo (Freeman et al. 2000) within scientific and social spaces, respectively. Figure 2.1.a suggests that while cardiac stem cell research like this paper gained massive early attention, this did not sustain, manifesting fragile, overhyped ideas that could not withstand broader scrutiny across the scientific community or application across science. This is contrasted in Figure 2.1.b with the case of cancer immunotherapy, where research gradually diffused to distant research groups and topics before garnering significant attention.

Beyond papers, we trace the diffusion trajectories of 28,504 unique subfields in biomedicine (Azoulay, Fons-Rosen, and Zivin 2019), encompassing nearly 1.9 million papers and more than 80 million citations. Our analysis reveals that limited diffusion of biomedical knowledge is systematically associated with an early rise and abrupt drop in popularity. The bottom panels of Figure 2.1 display the average trajectories of subfields by distinguishing those that experienced a sharp decline or collapse in scientific attention from those that did not by the end of 2019. Furthermore, our post-hoc analyses show that the likelihood of collapses of subfields is positively associated with the concentration of publications from superstar biomedical researchers, echoing aspects of the Dr. Anversa case.

Figure 2.1: Representation of Different Diffusion Levels and Contrasting Diffusion Trajectories



Note: Panels **a** and **b** illustrate 3D kernel density plots of diffusion indices and citations for PMID 11777997 (Cardiac Stem Cell) and PMID 11015443 (Cancer Immunotherapy) in scientific and social spaces, respectively. Publication years associated with each article are aligned to zero for comparison. Annual diffusion indices and citation counts are computed using a two-year rolling average. Panels **c** and **d** show kernel density plots based on average diffusion indices and citations, standardized within subfield ages. These plots contrast subfields that experienced collapse below the 0.5% threshold (blue) with those that did not (red), across scientific and social spaces respectively.

In this way, our work highlights that restricted diffusion in science can effectively capture socio-epistemic bubbles. Complementing citation dynamics with diffusion patterns enriches our

identification of robust insight in biomedical science, which can be readily improved by discounting bubbles and promoting convergent results sourced through social and topical diversity.

Methods

Manifold Representations of Social and Scientific Space

To assess the diffusion of ideas in science from biomedicine, we train two high-dimensional vector representations using neural embedding models (Le and Mikolov 2014) for publications cataloged in the PubMed Knowledge Graph (PKG) (Xu et al. 2020). The PKG provides 15,530,165 disambiguated author IDs and 481,497 unique combinations of Medical Subject Headings (MeSH) from 29,339 MeSH descriptors and 76 qualifiers, each assigned to 28,329,992 and 26,666,615 MEDLINE-indexed publications, respectively, by the end of 2019. Each document in the PubMed database is assigned a unique document identifier, PMID. The database also contains the publications to the publication reference records, which integrates PubMed's citation data, NIH's open citation collection, OpenCitations, and the Web of Science.

We specifically adapt the Doc2vec model (Le and Mikolov 2014), a variant of the Word2vec model (Mikolov et al. 2013), originally developed to produce dense vector representations for documents or paragraphs from the words that compose them. This approach has previously been extended to generate high-dimensional representational vectors geometrically proximate to the degree that entities frequently share neighbors, contexts (Mikolov et al. 2013; Kozlowski, Taddy, and Evans 2019; Garg et al. 2018), or are connected via social ties (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016).

We consider that a biomedical research article can be characterized by a list of: 1) MeSH terms and 2) research collaborators. Consequently, we build two separate representational vector spaces to capture "scientific space" and "social space", respectively. For training our vector representations, we utilize the Python Gensim package (Radim Rehurek 2010). We specifically use the Distributed Bag of Words (DBOW) model, analogous to the skip-gram model from the Word2vec framework, and simultaneously train the vector position of constituting elements (MeSH terms or author IDs) along with document vectors. This results in two spaces trained on 100-dimensional vector representations for PMIDs and their constituent elements. Training and validation procedures are detailed in the Appendix (Measuring Knowledge Diffusion Through Document Embedding Spaces).

Delineating Biomedical Subfields

Biomedical knowledge obtains influence when others recognize and build on it (Bourdieu 1975; Foster, Rzhetsky, and Evans 2015). In this work, we seek to understand the dynamics of diffusion and shifting attention at the level of biomedical subfields, which we define as a group of biomedical publications tightly related to a medically and biologically relevant research topic, identified through the PubMed Related Algorithm (PMRA) (Lin and Wilbur 2007). This method has been previously employed in studies examining the impact of publication retraction (Azoulay, Furman, and Murray 2015), repercussions of scientific scandal on careers (Azoulay, Bonatti, and Krieger 2017), shifts in research focus by scientists in response to NIH funding changes (Myers 2020), negative impacts from prize-winning on recipient competitors (Reschke, Azoulay, and Stuart 2018), and consequences of the premature death of elite life scientists (Azoulay, Fons-Rosen, and Zivin 2019) on subfields.

We specifically use the 28,504 unique seed articles curated by the Azoulay team. (2019), derived from publications by "superstar" biomedical scientists. Applying the PMRA-powered similar article function in PubMed enables us to capture over 1.9 million unique articles associated with these subfields published through 2019. We then extract ~86.8 million paper-to-paper citations identified by PKG based on them. A more comprehensive illustration of the original data source and our extension is available in the Appendix (Delineating Biomedical Subfield). To ensure robustness, we perform complementary analyses that redefine subfields based on the position of papers within our scientific embedding space, resulting in the same pattern of findings. Details and results are reported in the Appendix (Alternative Identification of Subfields).

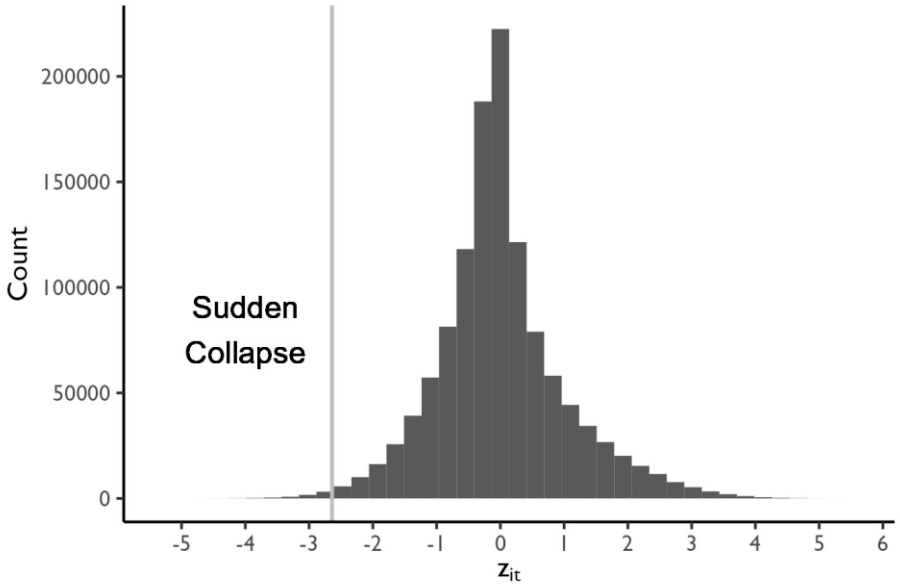
Outcome Event: Bubble bursting

Our primary outcome of interest is the event of socio-epistemic bubbles bursting, characterized by an abrupt decline in popularity of a given subfield that we measure in the decline of citation counts as illustrated in Figure 2.1. Specifically, we time bubble bursts based on when the standardized citation count difference of a given year from a subfield falls below extreme cutoffs within the life cycle of each subfield. This requires distinguishing subfields that experienced deflationary bursting, or collapse, from those that did not. We achieve this through the following steps.

We first compute $\Delta_{i,t} = c_i(t) - c_i(t-2)$, where $c_i(t)$ is the citations that a subfield i garnered during year t across 1970 to 2019. Unlike the approach taken by Azoulay et al. (2019) that uses publications indexed both in Web of Science and MEDLINE, we use all PMID to PMID citation links identified in PKG 2020 data to compute citation counts. (We include all

MEDLINE indexed publications, even when MeSH terms or author disambiguated IDs are not assigned to them.) Then, we standardize $\Delta_{i,t}$ within the life cycle of each subfield to make the $\Delta_{i,t}$ values comparable across 28,504 subfields. This is achieved by transforming $\Delta_{i,t}$ to $z_{i,t}$ by subtracting the mean of $\Delta_{i,t}$, $\overline{\Delta_i} = \frac{1}{N}\sum \Delta_{i,t}$, from $\Delta_{i,t}$ and dividing it by the standard deviation of $\Delta_{i,t}$ computed within a subfield. By doing so, we obtain the distribution of the standardized two-year citation difference, $z_{i,t}$, across 28,504 subfields. The distribution of $z_{i,t}$, with the range of [-5.2, 5.52], is presented in Figure 2.2.

Figure 2.2: Distribution of $z_{i,t}$ from 28,504 Subfields



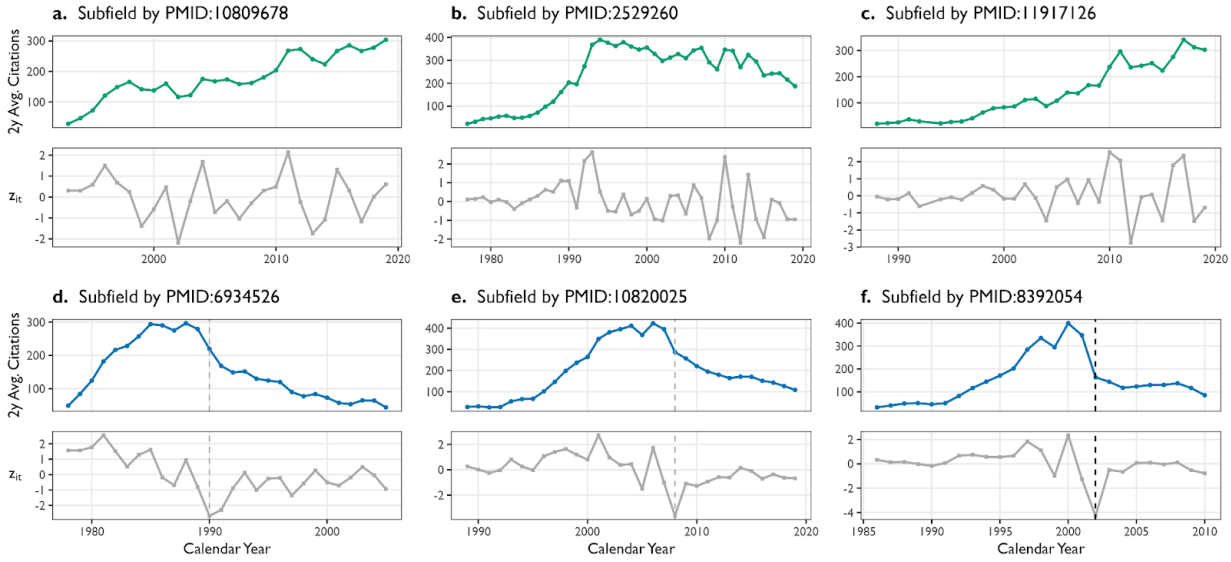
Note: The cutoff value for bubble burst here is set to -2.64, the bottom 0.5% percentile. The range of $z_{i,t}$ is [-5.2, 5.52].

We operationalize bubble bursts as when the standardized citation count difference for a given year in a subfield, $z_{i,t}$, falls below extreme cutoffs, such as 0.5%, 0.25%, or 0.1% of the distribution. To qualify a decline as a burst, we require that the average of $z_{i,t}$ after the drop must be negative, ensuring a continued loss of attention. Additionally, the peak citation count at the

subfield level should not occur in 2019, the final year of our dataset. If a subfield experiences more than one sharp decline, we consider the year with the most substantial one as the time of the burst. We note that bursts are preceded by bubbles: fields that experience these extreme drops also manifest greater than expected citations prior to collapse.

Using the 0.5% cutoff (i.e., $z_{i,t} < -2.64$) identifies 4,480 subfields (15.7% of 28,504 subfields) that experienced a sharp decline in collective scientific attention relative to other subfields. Applying the 0.25% ($z_{i,t} < -2.91$) and 0.1% ($z_{i,t} < -3.26$) cutoffs return 2,297 and 918 subfields with the bubble bursting events, respectively. Figure 2.3 contrasts three examples of subfields that did not experience these bubbles and bursts (top panels) with three examples that exhibited substantial declines in attention (bottom panels), according to our procedure described above.

Figure 2.3: Six Examples of Subfields



Note: Annual citation counts aggregated at the subfield level, using forward citations to related publications. Top panels (a, b, c): Subfields represented by three PMIDs, illustrating cases without bubble bursting events. Bottom panels (d, e, f): Subfields that experienced bubble bursting, corresponding to the cutoffs closest to the 0.5%, 0.25%, and 0.1% thresholds of $z_{i,t}$ value.

Key Indicator: Knowledge Diffusion

The key leading indicator for our analysis is subfield-level knowledge diffusion. We measure the knowledge diffusion by employing a 2-year rolling window approach. For each year, we identify papers published either in that year or the preceding year referencing at least one article published within a given subfield. We then separately calculate the average cosine distances (or 1-cosine similarity) between these focal articles in the subfield and the citing papers in our scientific and social spaces. This consideration leads us to measure two diffusion indices: 1) Diffusion across *Scientific* Space and 2) Diffusion across *Social* Space. In our model, we incorporate a one-year lag to assess the association between diffusion dynamics and the subsequent decline in citations.

Further Characterization of Subfield Dynamics

A. *Time Effect*

- ***Subfield Age.*** The difference between calendar years and the year seed articles were published is captured using subfield age dummies, included for each subfield up to the end of 2019. This approach controls for trends related to the age of the subfield without imposing a functional form.

B. *Subfield Growth Pattern*

- ***Cumulative Subfield Size.*** The total number of articles published in a subfield up to a given year. This measure controls for the potential impact of a subfield's size on citation dynamics. We apply a logarithmic transformation to address skewness for robust statistical comparisons between subfields of varying sizes.

- ***Two Rolling-Years Marginal Growth*.** The proportion of articles published in the current year and the previous year, divided by cumulative subfield size. This metric provides a normalized indicator of how actively a subfield is growing, shrinking, or remaining stagnant, adjusting for short-term fluctuations in publication activity that might affect outcomes of interest, such as citation dynamics.

C. Citation Dynamics

- ***Total Cumulative Citations***. Aggregate citation counts that publications within a subfield have received up until a specified year. We include this variable to control for the overall academic impact of a subfield, which may influence the likelihood of sudden changes in citation patterns. A natural logarithmic transformation is applied to address skewness.

- ***Two-year Rolling Citation Counts***. Citations a subfield accumulates during the given year and the past year. We take the natural logarithm of the raw counts. This variable controls for the recent volume of citations, separate from long-term trends.

- ***Gini Coefficient of Citation Counts***. The Gini coefficient measures the degree of centralization in citation counts within a subfield. The coefficient ranges from 0 (where every article in a subfield receives the same number of citations) to 1 (where a single article receives all citations). Annually, we compute the Gini coefficients for 1) *Total Cumulative Citations* and 2) *Two-year Rolling Citation* to control for the potential impact of citation concentration.

D. Other Controls

- ***Article Retraction Notification.*** Indicator variable that switches from 0 to 1 once a retraction notification is observed in a subfield. It controls for the potential impact that experiencing a retraction event in the subfield level might have on overall attention the subfield receives.

- ***After the Death (of Superstar Scientists)***. Indicator variable that switches from 0 to 1 with the death of superstar scientists (Supplementary Information 3.1). This attempts to capture any residual temporal effects of star death on citation dynamics.

- ***After Death (of Superstar Scientists) * Subfields Associated with Premature Death of Superstar Scientists.*** The first term is as previously described. The latter is an indicator variable that differentiates subfields associated with the premature deaths of elite scientists from those that are not. This controls for the impact of the sudden death of star scientists on citation dynamics, reflecting how the data set was originally constructed and the finding that star death is *positively* associated with increases in subfield citation (Azoulay, Fons-Rosen, and Zivin 2019).

- ***Calendar Year Fixed-Effect.*** Year dummies to account for potential effects of the calendar year from 1970 and 2019. We include this to ensure that any time-specific external influences are controlled across all subfields.

- ***Strata ID***. 3,076 "strata" IDs identified from the subfields associated with publications of

prematurely deceased superstars (Azoulay, Fons-Rosen, and Zivin 2019). These IDs are assigned to comparable "within strata" subfields not experiencing a loss of star scientists. These comparable subfields are matched with those experiencing a star death based on key metrics such as 1) publication years, 2) team sizes, 3) ages of associated scientists, and 4) long-run citation impact.

Model

Using a nonparametric Cox model and discrete-time event history model, we relate the annual diffusion indices for each subfield calculated across social and scientific spaces with an abrupt decline in the relevance of a given subfield, or "bubble burst," as illustrated in Figure 2.1.a. Formally, the discrete-time event history analysis model can be written as:

$$log(\frac{p_{ti}}{1-p_{ti}}) = \alpha D_{ti} + \beta x_{(t-1)i} \quad \cdots \text{Eq. (2.1).}$$

p_{ti} denotes the probability of event happens at t for subfield i,
D_{ti} denotes time dummies corresponding to t with coefficients α,
x_{ti} is vector for covariates (time varying and constant over time) with coefficients β.

Results

Contrasting Trajectories of Cardiac Stem Cell Research and Cancer Immunotherapy

Applying neural embedding models to MEDLINE data enables us to project all biomedical research articles onto scientific and social manifolds. As detailed in Methods and Supplementary Information (S2), this allows us to locate their relative positions within collaborative networks of scientists and biomedical entities through research. The cosine or angular distances between citing and cited research measured over social and scientific spaces aggregate into straightforward, continuous metrics of diffusion. To demonstrate the effectiveness

of our approach utilizing these scientific and social spaces, we examine trajectories of two highly cited publications at the individual paper level, each drawn from Cardiac Stem Cell and Cancer Immunotherapy research, respectively.

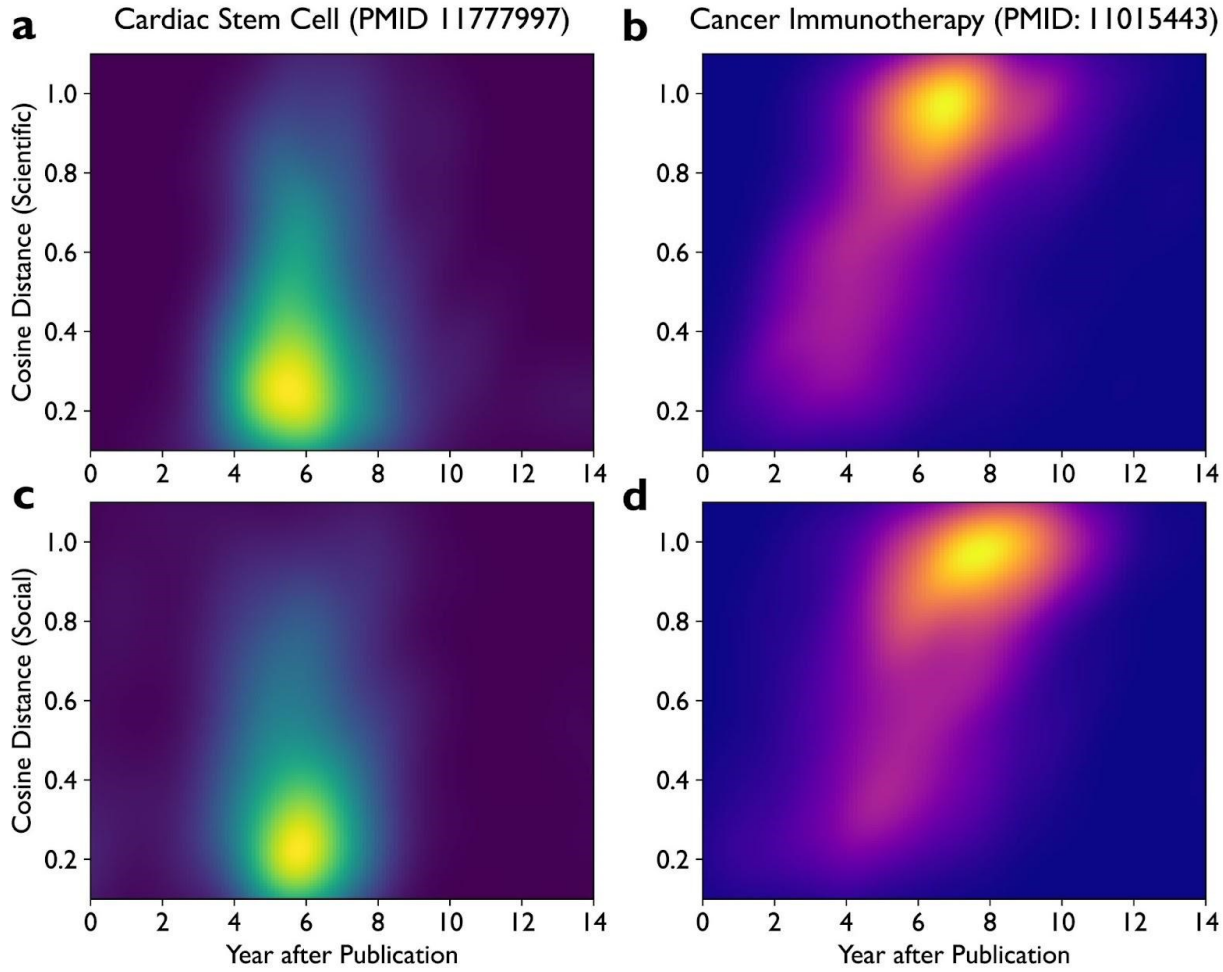
Our first case is a research article published (PMID: 11777997) in *the New England Journal of Medicine* in 2002 (Quaini et al. 2002). Led by Dr. Piero Anversa, this research supported the existence of substantial numbers of endogenous myocardial stem and progenitor cells, proposing their potential to regenerate heart muscle. This line of research initially received outsized attention because it suggested new possibilities for heart regeneration after severe myocardial infarctions involving massive tissue loss. This claim was later called into question by several researchers outside the Anversa network, however, eventually leading to the retraction of more than 30 papers by 2018 from claims of data fabrication and scientific malpractice(Chien et al. 2019).

Conversely, the second example, an article (PMID: 11015443) published in *the Journal of Experimental Medicine* in 2000 (Freeman et al. 2000) represents a study by a team of pioneering researchers in the field of cancer immunotherapy. Their work focuses on the inhibition of negative immune regulation and its implications for cancer treatment. The publication and subsequent work spurred the development of a broad spectrum of cancer immunology and immunotherapy research initiatives across many research groups and countries globally, laying the groundwork for what has become one of the most impactful innovations in cancer treatment.

The upper panels (**a** and **b**) of Figure 2.1 visualize the contrasting temporal trajectories of these two publications in size of attention and diffusion within the scientific and social space, respectively, with 3D kernel density estimation. Figure 2.4 projects the estimated density onto

2D heatmaps. Annual diffusion indices are computed using citation data from the given and previous year. This rolling two-year window averages cosine distances between the focal and forward citing papers across social and scientific space, providing a dynamic measure of diffusion over time.

Figure 2.4: 2D Heatmaps for the Upper Panels of Figure 2.1



Note: Values are derived from the identical kernel density estimations graphed in Figure 2.1 for the distribution of diffusion indices in scientific (panels **a** and **b**) and social spaces (panels **c** and **d**), respectively.

Dr. Anversa's publication experienced a meteoric rise in total citations during the first five years following debut. However, our measures indicate limited diffusion across the scientific space of distinct subfields and the social space of author teams citing the paper, which preceded

a sharp decline in attention toward the paper, resembling the burst of market bubbles. In contrast, the article on Cancer Immunotherapy, which demonstrated the potential to inhibit negative immune regulation in treating cancer, gained early attention at a much slower pace. Nevertheless, the ideas ultimately diffused much more broadly, becoming one of the most influential innovations in recent cancer treatment and research. This culminated in awarding the 2018 Physiology and Medicine Nobel Prize to Drs. Tasuku Honjo and James P. Allison for advancing the scientific understanding of Cancer Immunotherapy. These contrasting cases demonstrate how our diffusion metric accounting for epistemic bubbles offers a more nuanced understanding of scientific influence than traditional citation counts, capturing the complex dynamics of diffusion through social and scientific spaces and its potential consequences.

Knowledge Concentration Anticipates Collapse

We elevate our analysis to the level of scientific subfields to systematically test the generalizability of our approach. We apply our framework to 28,504 unique biomedical subfields curated by Azoulay et al. (2019). Each subfield encompasses a compactly defined set of biomedical research articles using the PubMed Related Article (PMRA) algorithm (Lin and Wilbur 2007) applied to a given seed article. This algorithm underpins the official PubMed interface, serving as a pivotal tool for researchers to locate articles related to a focal research paper, which has been fruitfully used in various studies, such as repercussions of scientific scandal on careers (Azoulay, Bonatti, and Krieger 2017), shifts in research focus among scientists responding to NIH funding changes (Myers 2020), and the negative impact of winning prizes for recipient competitors (Reschke, Azoulay, and Stuart 2018). The subfields this approach allows us to identify enable us to analyze diffusion dynamics, epistemic bubbles, and

collapses of scientific attention beyond selective, high-profile papers. Specifically, if work from a focal subfield is predominantly cited by research in close social and scientific proximity, the subfield's insights may not diffuse despite its seeming popularity and could retain inflated value due to local reinforcement. In other words, we anticipate that substantial and dramatic declines in the popularity of research ideas, conceptualized as knowledge 'bubbles bursting,' can be predicted by the degree to which these ideas, despite their apparent popularity, have failed to diffuse across the social and scientific space via citations.

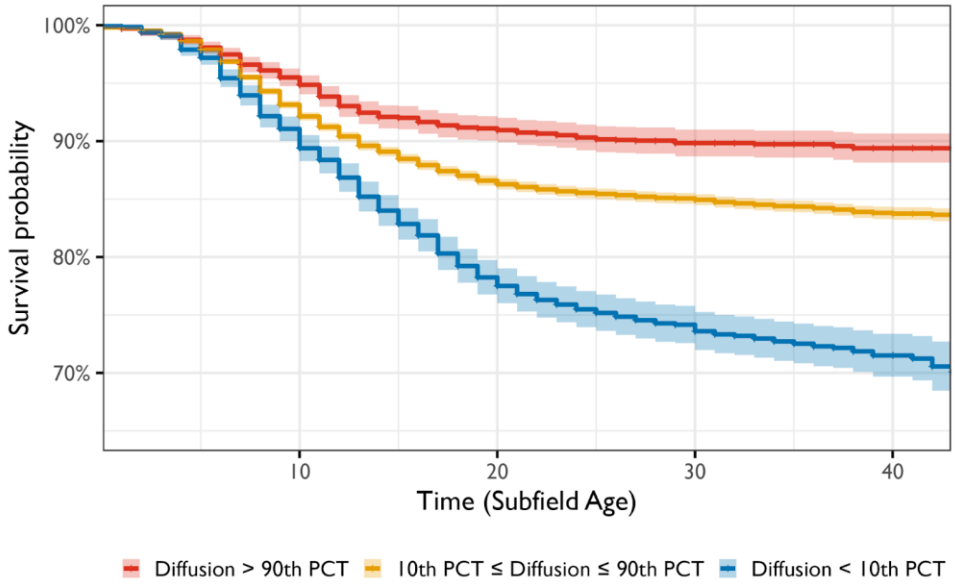
Our primary outcome of interest is 'bubble bursting' or collapse, defined as an abrupt decline in the relevance of a given subfield for science. We time a bubble burst by comparing the standardized citation difference that a subfield garners in a given year to its performance two years prior, marking if it falls below an extreme threshold. This approach allows us to distinguish subfields that experienced deflationary bursts from those that did not by basing each standardized citation count difference against the values derived from 28,504 unique subfields. We use the bottom 0.5% of the distribution of standardized citation differences as our threshold, which captures 4,480 out of 28,504 unique subfields as experiencing a collapse. To ensure the robustness of our results, we also apply thresholds of 0.25% and 0.1%, identifying 2,297 and 918 collapsed subfields respectively, and report the results from parallel analyses using these thresholds throughout the following analyses and in the Supplementary Information. Across these operationalizations, the subfields that experience a collapse also experienced a significant positive deviation from expected citation rates preceding collapse (Table 2.5). Fields that experience a disproportionate deflation experienced a previous inflation. In short, bubbles burst.

We compute our knowledge diffusion indices, our main predictors, for each subfield across scientific and social spaces. We identify papers published that reference at least one

article within each subfield. We then calculate the average cosine distances between the referenced articles in each subfield and the citing papers with 2y rolling windows, separately for scientific and social spaces to measure scientific and social diffusion (Methods).

Using a nonparametric Cox survival model to predict the probability of bubble bursting, our estimation reveals the knowledge diffusion index as a strong leading signal preceding a sudden collapse in attention. We employ a one-year lag for our diffusion measures when associating them with the outcome of interest, collapse of attention.

Figure 2.5: Survival Probability against Bubble Bursting as a Function of Knowledge Diffusion in Social Space



Note: Events are defined as a sharp decline of 2-year citation counts at the subfield level with 0.5% cutoff (Method). Survival refers to the converse, i.e., not experiencing a subfield-level extreme deflationary event. Subfield ages are set to 0 in the year when the focal seed article spanning a subfield was published. Diffusion percentile is ranked within calendar years and subfield ages. Bands depict 95% confidence intervals.

By splitting our observations into three groups with diffusion percentiles ranked by calendar year and subfield age—the bottom 10th percentile, the top 10th percentile, and the middle between them—Figure 2.5 visualizes that diffusion in the social space forecasts the bursting of attention bubbles captured by the 0.5% threshold. The result indicates that low

diffusion rates may signal poor long-term subfield survival. Conversely, high diffusion is related to subfield survival in the long term, avoiding extreme subfield-level deflationary events. We confirm this pattern, presented in Figure 2.5, with discrete-time event history models that allow us to consider temporal covariates, including field size and growth rate, total cumulative citations, citation concentration across papers, paper retractions, and unexpected deaths of elite scientists.

Our analysis consistently shows that the lower a paper's diffusion of influence, the greater the hazard that the subfield will experience an abrupt collapse of attention (Table 2.1 and Table A.2.1). For example, as diffusion in social space reduces from one standard deviation above to one below the mean, it translates into a 74.02% (95% CI: 43.61%–110.85%) increase in the odds of experiencing a major reduction in scientific attention, accounting for subfield age, calendar year, and other covariates. Tables A2.2 and A2.3 show the estimations based on 0.25% and 0.1% thresholds to identify burst subfields.

Overall, we observe a more pronounced impact of limited social diffusion than scientific diffusion on the likelihood of subfield collapse. We posit that this likely stems from tacit confounders in research that emerge when conducted by a concentrated, connected group of scientists. When close-knit groups perform research under uniform assumptions, methodologies, and even shared resources, their findings are less likely to replicate among outsiders (Danchev, Rzhetsky, and Evans 2019; Belikov, Rzhetsky, and Evans 2022). By contrast, the applicability of verified scientific findings across different biomedical domains may vary. A therapy's effectiveness for treating breast cancer is undiminished by its irrelevance for heart disease. But the failure of findings to diffuse across different groups of scientists in the same area indicates a limitation of their published scientific knowledge.

Table 2.1: Model Estimates with the Bottom 0.5% Cutoff for Citation Differences in the Two-Year Rolling Period

Dependent Variable	***Substantial Decline of Citations***				
	Estimate	**Std. Error**	***t***	***p*-value**	**95% C.I.**
Diffusion					
Scientific Space	-0.204	0.039	-5.228	< 0.001	[-0.280, -0.128]
Social Space	-0.277	0.049	-5.598	< 0.001	[-0.373, -0.181]
Subfield Growth Pattern					
Cum. Subfield Size (logged)	0.499	0.084	5.923	< 0.001	[0.334, 0.664]
2-year Subfield Marginal Growth	-0.217	0.010	-21.543	< 0.001	[-0.237, -0.197]
Citation Dynamics					
Cum. Citations (logged)	-2.472	0.164	-15.003	< 0.001	[-2.793, -2.151]
2-year Citations (logged)	2.772	0.146	18.942	< 0.001	[2.486, 3.058]
Gini Coef. of Cum. Citations	0.012	0.006	1.843	0.065	[0.000, 0.024]
Gini Coef. of 2-year Citations	-0.026	0.006	-4.710	0.001	[-0.038, -0.014]
Other Controls					
Retraction Notice Published	0.062	0.199	0.313	0.754	[-0.328, 0.452]
After Death	0.152	0.129	1.186	0.236	[-0.101, 0.405]
After Death * Superstar Death	-0.138	0.086	-1.601	0.109	[-0.307, 0.031]
Log-Likelihood			-26,289.5		
Total Observations			1,313,433		

*Note***:** Coefficients for fixed effects of field age, calendar year, and strata ID dummies are omitted. Variables under *Knowledge Diffusion*, *Subfield Growth Pattern*, *and Citation Dynamics* are all one-year lagged. The diffusion indices are standardized within field ages and calendar years across 28,504 subfields. Standard errors are clustered with strata ID and calendar years.

Post-Hoc Analysis

To gain further insights into the phenomenon of scientific bubbles, we conduct a series of subsequent analyses to gain deeper insight into characteristics of socio-epistemic bubbles and consequences of their collapse with augmented data, including subfield characteristics provided by the Azoulay team (2019) and information extracted from the NIH's iCite system (Hutchins et al. 2019, 2016).

Stars' Importance

In exploring the mechanisms associated with the likelihood of scientific bubbles bursting, as suggested by the Stem Cell Cardiac regeneration case, we examine the relationship between 'Star Importance to the Subfield' and the likelihood of subfield collapses. To do so, we draw on the replication data provided by the Azoulay team, defining 'Star Importance to the Subfield' as the fraction of papers authored by superstar scientists within the subfield (variable name: 'imprtnc'). Utilizing logistic regression, we assess whether this measure can predict the likelihood of collapses versus uncollapsed subfields. As detailed in the Appendix (Delineating Biomedical Subfield**)**, while the original dataset contains 28,504 unique seed articles, it yields 34,218 pairs of subfield strata and seed articles for subfield identification. Thus, we applied clustered standard errors at the Strata IDs.

Table 2.2: Star's Importance to the Subfield and Collapse

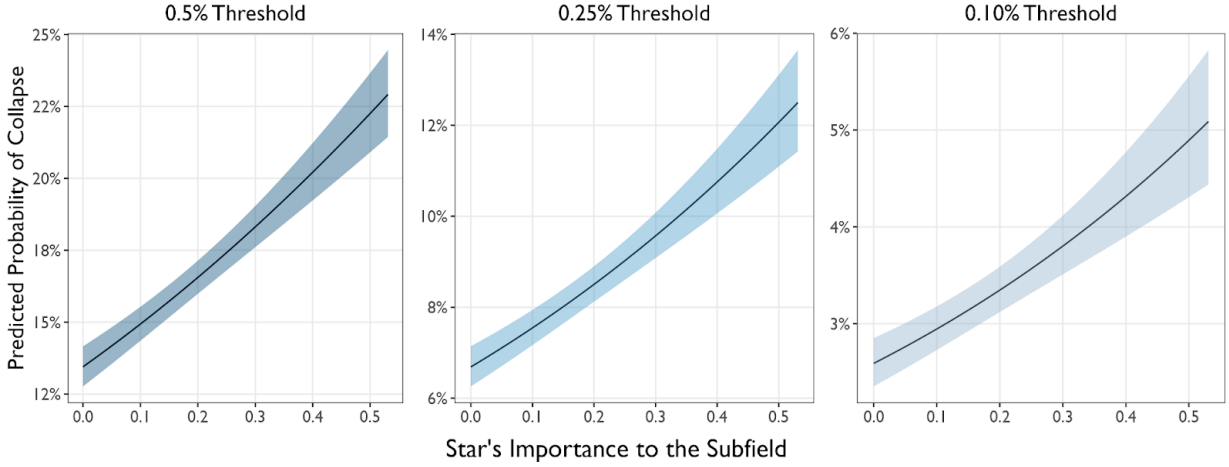
Dependent Variable	***Collapsed versus Not Collapsed***		
Threshold	**0.5%**	**0.25%**	**0.1%**
(Intercept)	-1.862*** [-1.922, -1.802] ($p < 0.001$)	-2.636*** [-2.706, -2.565] ($p < 0.001$)	-3.627*** [-3.725, -3.528] ($p < 0.001$)
Star's Importance to the Subfield	1.222*** [1.001, 1.443] ($p < 0.001$)	1.299*** [1.038, 1.560] ($p < 0.001$)	1.320*** [0.961, 1.680] ($p < 0.001$)
Log-Likelihood	-14,899.2	-9,588.9	-4,814.2
Total Observations (# of Unique Seed Article-Strata Pairs)		34,218	

*Note***:** The 95% confidence intervals and p-values are based on the standard errors clustered at strata ID.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table 2.2 shows a significant association between the stars' importance and the likelihood of collapse; Figure 2.6 visualizes the estimation reported in Table 2.2, which suggests the association between concentration of scientific capital and the probability of a subfield collapsing.

Our analysis revealed that the importance of superstar scientists within a subfield, as quantified by the proportion of their publications per subfield, is positively correlated with the likelihood of collapse compared to subfields that did not burst. "Star scientists" whose work dominates a subfield (Azoulay, Fons-Rosen, and Zivin 2019) are more likely to have their early findings overhyped, and subsequently "burst", than subfields without a star.

Figure 2.6: Predicted Probability of Collapse by the Star's Importance to the Subfield, Based on the Estimates in Table 2.2



Note: The range of Star's Importance to the Subfield extends up to 3 standard deviations from the distribution of the variable. The mean and standard deviation of the variable, Star's Importance to the Subfield, are 0.151 and 0.127, respectively. The bands represent the 95% confidence intervals, calculated based on the standard errors reported in Table 2.2.

Subfield Funding Accounted by Star's Collaborators

We additionally examine the relationship between the 'Fraction of Subfield Funding Accounted by Star's Collaborators' and the likelihood of subfield collapses, using the same approach used for 'Star's Importance.' This analysis assesses the association between the 'fraction of subfield funding accounted for by collaborators' (variable name: 'frac_collabs_field_nih_fndg') and collapses. Table 2.3 demonstrates a significant association between the concentration of funding among star scientists' collaborators and the collapse.

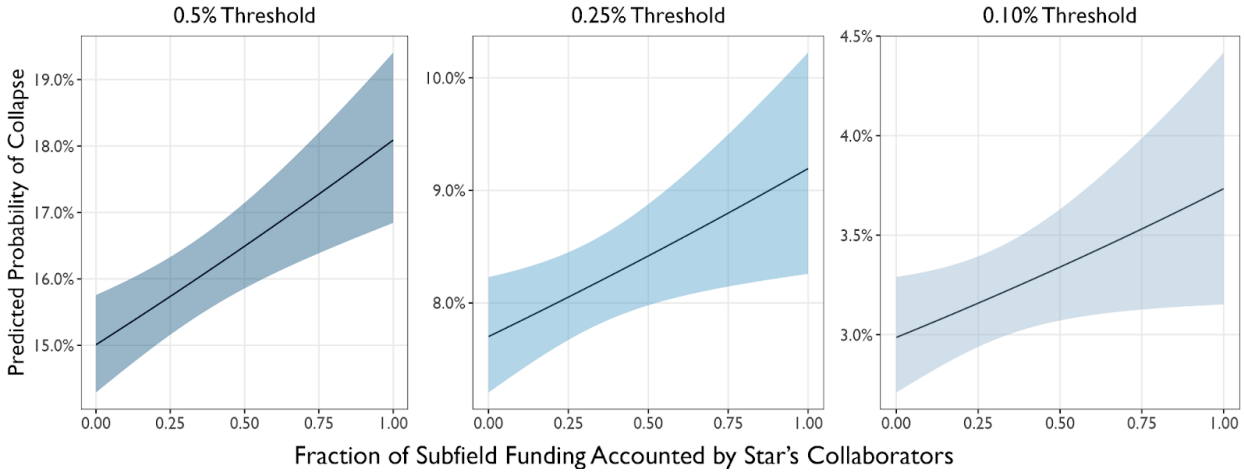
Table 2.3: Fraction of Subfield Funding Accounted by Star's Collaborators

Dependent Variable	***Collapsed versus Not Collapsed***		
Threshold	**0.5%**	**0.25%**	**0.1%**
(Intercept)	-1.734*** [-1.791, -1.676] ($p < 0.001$)	-2.484*** [-2.556, -2.412] ($p < 0.001$)	-3.481*** [-3.581, -3.381] ($p < 0.001$)
Fraction of Subfield Funding Accounted by Star's Collaborators	0.224*** [0.110, 0.337] ($p < 0.001$)	0.194* [0.037, 0.350] ($p = 0.015$)	0.231* [0.004, 0.458] ($p = 0.046$)
Log-Likelihood	-14,899.2	-9,588.9	-4,814.2
Total Observations (# of Unique Seed Article-Strata Pairs)		34,218	

Note: The 95% confidence intervals and p-values are based on the standard errors clustered at strata ID.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Figure 2.7 visualizes the estimation reported in Table 2.3, further suggesting an association between the concentration of scientific capital and the probability of a subfield collapsing.

Figure 2.7: Predicted Probability of Collapse by Fraction of Subfield Funding Accounted by Star's Collaborators, Based on the Estimates in Table 2.3



Note: The mean and standard deviation of the variable, Fraction of Subfield Funding Accounted by Star's Collaborators, are 0.283 and 0.293, respectively. The bands represent the 95% confidence intervals, calculated based on the standard errors reported in Table 2.3.

We find a positive association between the fraction of NIH funding allocated to collaborators of these star scientists and the likelihood of attentional collapse. This suggests that limited diffusion and subsequent collapses may be correlated with the concentration of

"scientific capital" in terms of reputation and resources (Bourdieu 1975), as exemplified by the Stem Cell Cardiac case discussed above.

Approximate Potential for Clinical Translation

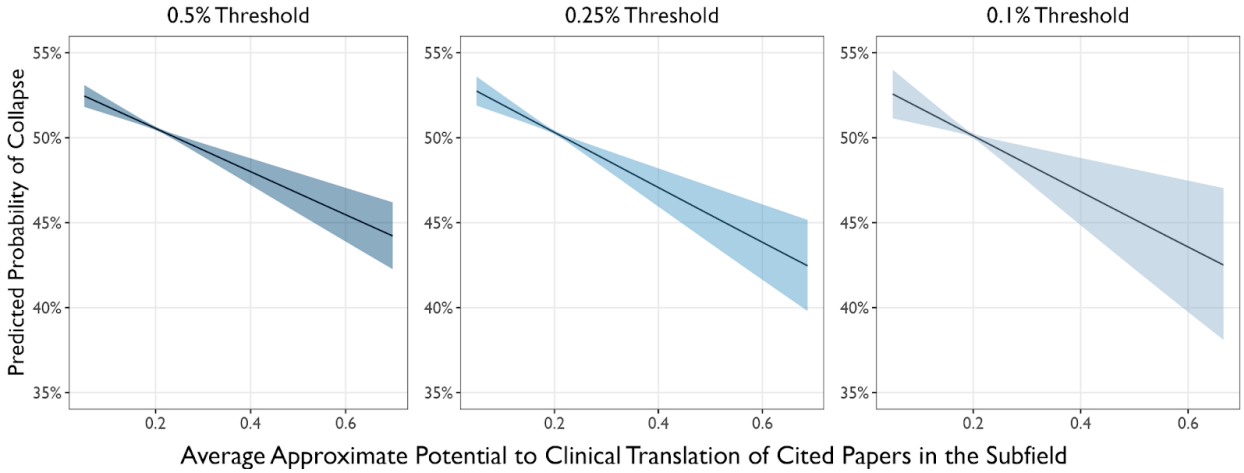
We further augmented our dataset with the Approximate Potential for Clinical Translation (APT) modules (Hutchins et al. 2019) from PubMed's iCite system. This module evaluates the likelihood of a paper's research being applied and cited in subsequent clinical studies. We extracted APT values for each publication in our dataset from the iCite bulk data. These values were then aggregated at the subfield level by calculating the average APT for articles cited at least once up to a specific calendar year. We selected the year when collapsed subfields experienced collapses and matched this time to the calendar year (and consequently the field age) of subfields that did not collapse within the strata initially established by the Azoulay team. The subsequent logistic regression, using average APT values aggregated by calendar year and field age as predictors, indicates that a higher potential for clinical translation may prevent a drastic collapse, as Table 2.4 and Figure 2.8 report.

Table 2.4: Approximate Potential to Clinical Translation

Dependent Variable	***Collapsed versus Not Collapsed***		
Threshold	**0.5%**	**0.25%**	**0.1%**
(Intercept)	0.124*** [0.09, 0.158] ($p < 0.001$	0.142*** [0.097, 0.188] ($p < 0.001$)	0.136*** [0.059, 0.213] ($p < 0.001$)
Approximate Potential to Clinical Translation	-0.509*** [-0.672, -0.347] ($p < 0.001$)	-0.650*** [-0.875, -0.424] ($p < 0.001$)	-0.658*** [-1.047, -0.268] ($p < 0.001$)
Log-Likelihood	-6,073.3	-3,348.3	-1,365.8
Total Observations (# of Unique Seed Article-Strata Pairs)	8,773	4,840	1,974

Note: The 95% confidence intervals and p-values are based on the standard errors clustered at strata ID-Field age pair. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Figure 2.8: Predicted Probability of Collapse by the Average Approximate Potential to Clinical Translation of Cited Papers within the Subfield, Based on the Estimates in Table 2.4.



Note: The range of Average Approximate Potential to Clinical Translation of Cited Papers within the Subfield extends up to 3 standard deviations from the distribution. The bands represent the 95% confidence intervals, calculated based on the standard errors reported in Table 2.4.

Comparing Actual vs. Expected Citations in Collapsed and Uncollapsed Subfields.

To further support the concept of bubble as 'inflated' attention and our operationalization of it, we investigate how the citation counts that a subfield garners before collapse deviate from expected citations, comparing these discrepancies between collapsed and uncollapsed subfields.

We utilized the NIH's iCite Influence Module, which provides an annual 'expected citation count' for each MEDLINE-indexed paper. This system offers a benchmark for the number of citations a typical (median) MEDLINE-indexed paper, identified from the co-citation network and published in the same year, would receive (Hutchins et al. 2016). After extracting the expected annual citation count for all publications in our dataset, we summed these expected citation counts for each subfield across each calendar year for the articles published up to those years.

This approach enabled us to compute the difference between the actual citations a subfield garnered and the expected count. We then focused on the one and two years before the collapse of subfields. We matched these years with subfields that did not collapse within the

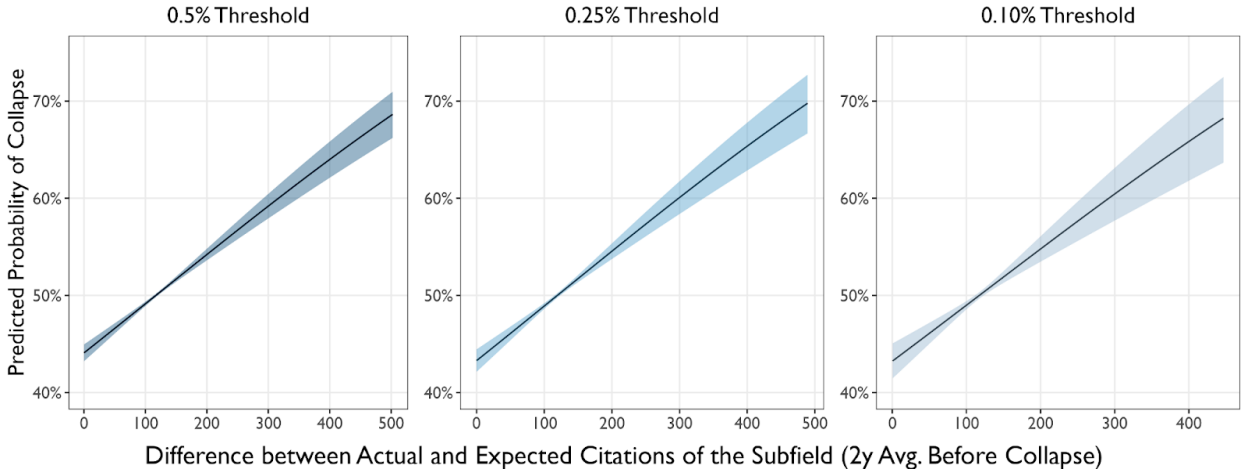
same strata (i.e., same publication year, similar long-term citation counts, similar star-scientist age, distinct intellectual location), analogous to the previous analysis for clinical translation.

Using logistic regressions, with average deviation from expected citation counts aggregated by calendar year and strata to capture the degree of "inflation" as a predictor, the analysis presented below in Table 2.5 and Figure 2.9 suggests that a greater degree of deviation from expected citation counts is positively associated with the likelihood of a subfield experiencing a collapse in the subsequent year. This pattern represents a positive indication of the attention bubble that may subsequently burst. When combined with results from the main analysis, this indicates that positive deviations, or the indication of bubbles, are inversely correlated with diffusion.

Table 2.5: Difference between Actual and Expected Citations before Collapse (2y)

Dependent Variable	***Collapsed versus Not Collapsed***		
Threshold	**0.5%**	**0.25%**	**0.1%**
(Intercept)	-0.238*** [-0.273, -0.203] ($p < 0.001$)	-0.270*** [-0.317, -0.223] ($p < 0.001$)	-0.273*** [-0.346, -0.199] ($p < 0.001$)
Difference between Actual and Expected Citations Before Collapse (2y)	0.002*** [0.002, 0.002] ($p < 0.001$)	0.002*** [0.002, 0.003] ($p < 0.001$)	0.002*** [0.002, 0.003] ($p < 0.001$)
Log-Likelihood	-6,016.7	-3,314.1	-1,352.8
Total Observations (Matched Year-Field Age-Strata)	8,773	4,840	1,974

Figure 2.9: Predicted Probability of Collapse by the Difference Between Actual and Expected Citation, Based on the Estimates in Table 2.5



Note: The range of Difference Between Actual and Expected Citation, extends up to 3 standard deviations from the distribution. The bands represent the 95% confidence intervals, calculated based on the standard errors reported in Table 2.5.

Implications of Bubble Burst

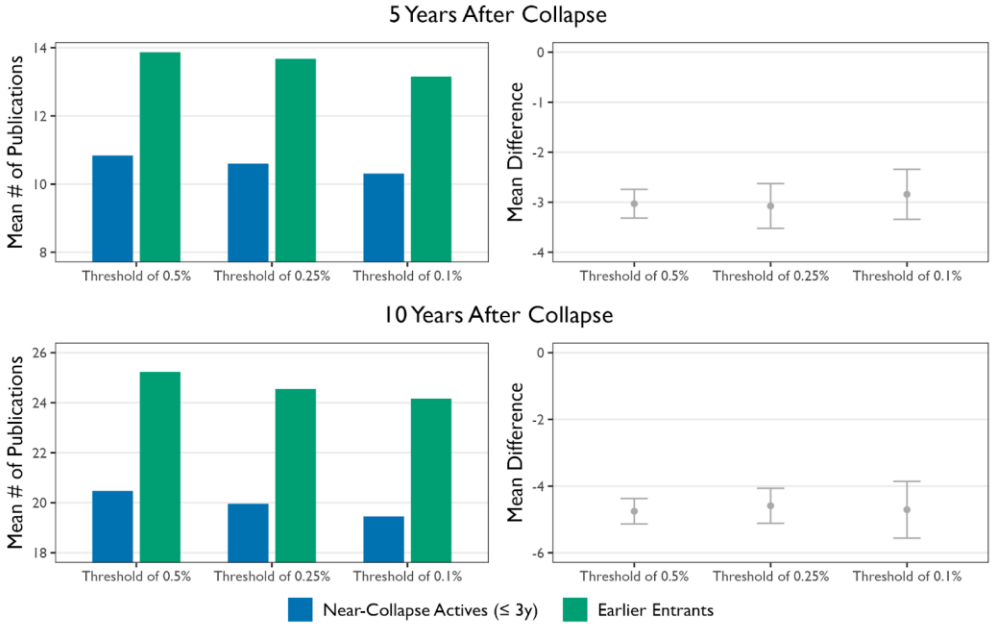
To evaluate the implications of epistemic bubbles and bursts, we compare the productivity of authors who published their articles close to the time of collapse (e.g., authors who published in 2001, 2002, or 2003 when the collapse was measured in 2003) with those who published in the same subfield at an earlier time (e.g., in or before 2000). As shown in Table 2.6 and Figure 2.10, findings suggest that those who entered right before collapse were significantly less productive in the mean number of publications both 5 and 10 years after collapse, compared to early entrants. This suggests that subfield collapse may shape researchers' reputations and career outcomes.

Table 2.6: Pairwise *t*-test Comparing Average Productivity Differences Between Near-Collapse Actives (≤2 Years Before Collapse) and Early Entrants

	Threshold	Estimate	t	*p*-value (*df*)	95% C.I.
5 Years	0.5%	-3.030	-20.67	< 0.001 (3,910)	[-3.317, -2.743]
	0.25%	-3.075	-13.48	< 0.001 (1,983)	[-3.522, -2.628]
	0.1%	-2.841	-11.13	< 0.001 (773)	[-3.342, -2.340]
10 Years	0.5%	-4.754	-24.45	< 0.001 (3,605)	[-5.136, -4.373]
	0.25%	-4.590	-17.17	< 0.001 (1,818)	[-5.114, -4.066]
	0.1%	-4.707	-10.87	< 0.001 (711)	[-5.558, -3.857]

Note: Subfields that collapsed after 2015 were excluded from the 5-year productivity comparison. Likewise, for the 10-year productivity, only subfields that collapsed on or before 2011 were included, considering the observation windows.

Figure 2.10: Comparing Author Productivity in Collapsed Subfields 5- and 10-Years Post-Collapse



Note: The error bars represent the 95% confidence intervals for the mean differences in average publication numbers. Comparisons are drawn between authors who entered the field early and those active near the collapse, based on paired t-tests.

We also consider the implications of bubbles for the allocation of research funding. We trace the average number of new grants acknowledged per year in papers across subfields. Our analysis shows that more than 80% of the subfields, which experienced a substantial decrease in

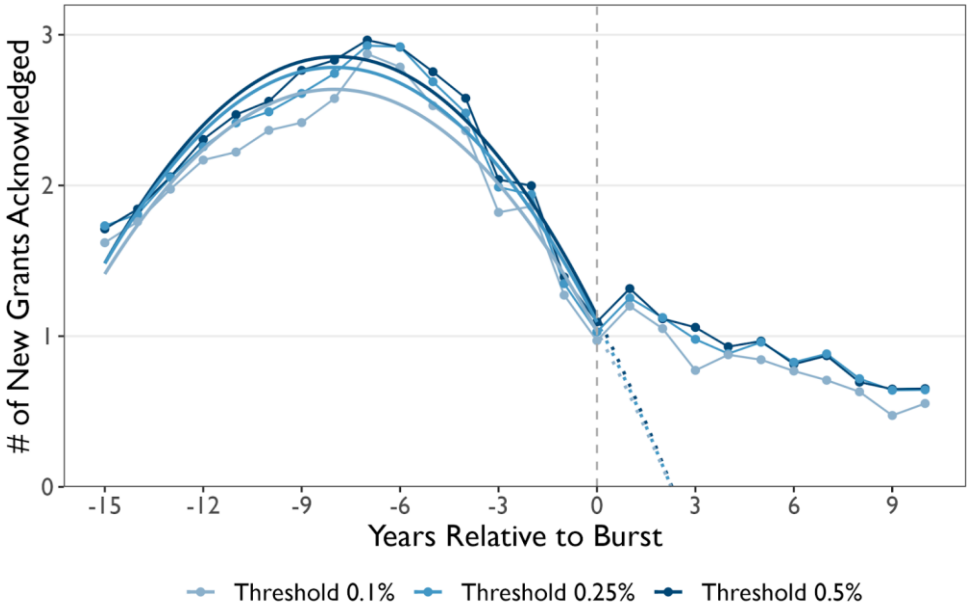
scientific attention, acknowledged new grants after collapse. By the end of 2019, the median number of such grants was 6, as detailed in Table 2.7.

Table 2.7: Proportion of Subfields with Newly Acknowledged Grants After Collapse, and the Mean, 1st Quantile, Median, and 3rd Quartile of the Number of New Grants Post-Collapse

Threshold	% of Subfields with New Grants Acknowledged After Collapse	Mean	Q1	Median	Q3
0.5%	83.12%	11.8	2	6	16
0.25%	82.93%	11.3	2	6	15
0.1%	81.70%	10.1	1	6	13

Note: Subfields that collapsed after 2015 were excluded from the 5-year productivity comparison. For the 10-year productivity analysis, only subfields that collapsed on or before 2011 were included, in consideration of the observation windows.

Figure 2.11: The Average Number of New Grants Acknowledged in Collapsed Subfields by Years Relative to Burst



Note: The quadratic fit is applied to data from years -15 to 0 relative to the burst year, with dotted lines representing extrapolations starting from year 0 onwards.

Figure 2.11 illustrates the trends from 15 years before to 10 years after the collapse. It shows that while peaks of new funding precede collapse, the rate at which funding decreases

after a burst is markedly slower than the trend observed before it. This pattern suggests a substantial lag by which money continues to support research that the broader biomedical community may perceive as less scientifically and clinically relevant.

Discussion

Current metrics of scientific attention and confidence pay scant attention to patterns of research consumption and diffusion across diverse people, institutions, disciplines, regions, and beyond. This lack of consideration can lead to an incomplete understanding of a research field's true impact and potential. Our knowledge diffusion index contrasts with and complements citation counts, the conventional unit of scientific credit. Citations alone are blind to who, where, and how far across the landscape of science those building on research reside, but our diffusion index provides a more comprehensive view.

A constriction in diffusion identifies an epistemic bubble or echo chamber that represents a leading indicator of future collapse in relevance and attention accorded to scientific and biomedical knowledge. Researchers can anticipate the collapse of biomedical approaches years prior to their occurrence by systematically tracking the diffusion of their ideas across scientists and biomedical areas. Additionally, science and biomedical policy that analyzes knowledge diffusion patterns can anticipate such collapses and may reduce their occurrence by incentivizing and accounting for diverse, disconnected support for robust scientific and medical claims (Belikov, Rzhetsky, and Evans 2022).

Like other methods aimed at quantitatively evaluating research impact, our framework for measuring diffusion and its implementation should not replace the holistic judgment of research quality. Furthermore, while we draw on the concept of 'bubbles' in science, analogous

to those in financial markets, it is worthwhile to recognize their unique aspects in the context of science. For example, small, dense research networks may be crucial for initiating high-risk projects at early stages despite a high probability of failure. In addition, scientific bubbles may not always arise from speculation, but could result from authentic scientific enthusiasm or localized beliefs in a promising research direction.

Nevertheless, our finding holds strong implications for biomedical researchers, science-based industries, and science policymakers. By accounting for diffusion and diversity, funding agencies can spot bubbles and adjust resource allocation by diversifying groups of researchers sponsored for a particular research topic. Research information platforms like PubMed, OpenAlex, the Web of Science, or Google Scholar could also incorporate strong, leading signals from which analysts can anticipate the future relevance of current research. A high diffusion index indicates that trending insights are more likely robust than fragile. Regular self-assessments of knowledge diffusion could enable individual researchers, teams, and labs to better gauge the robustness and future impact of their work. Further, documenting associations between scientific knowledge diffusion and its applications, as in the translation of biomedical research from bench to clinic, can better inform science policy.

Our results draw on subfields identified in academic science using a particular delineation of research subfields. Nevertheless, our analysis demonstrates clear evidence for the wisdom of diverse crowds in science and technology to sustain advance. It underscores the importance of both social and scientific diversity for robust evaluation of an idea's relevance to science as a whole. Moreover, our proposed framework for measuring diffusion may extend to other domains of knowledge, such as the spread of misinformation, by allowing us to measure diversity in information consumption (Kim et al. 2023). In social media, algorithmic metrics that account for

diversity in diffusion would be far less susceptible to strategic, concentrated efforts seeking to misclassify information as a legitimate, widespread trend (e.g., on Facebook's Newsfeed), just as they would decrease the intentional or unintentional illusion of scientific support.

In this way, we demonstrate the importance of idea diffusion for advancing scientific knowledge, its ability to transfer across broad science communities, and the relevance of these signals for forecasting robust ideas upon which to build novel and critical scientific and biomedical knowledge. Ultimately, our analysis underscores the relative importance of identifying the path of an idea's consumption over its point of production for predicting lasting, far-reaching impact. Accounting for this will enable the design of wise and diverse research, development, and clinical crowds, leading to improved research policy, greater reproducibility, and more sustained impact on future knowledge.

Appendix for Chapter 2

Table A2.1: Model Estimates with the Bottom 0.5% Cutoff, with and without Controls

Dependent Variable	*Substantial Decline of Citation*				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Knowledge Diffusion					
Scientific Space	-0.217*** [-0.293, -0.141] ($p < 0.001$)	-0.271*** [-0.350, -0.192] ($p < 0.001$)	-0.137*** [-0.207, -0.068] ($p < 0.001$)	-0.138*** [-0.208, -0.069] ($p < 0.001$)	-0.204*** [-0.28, -0.127] ($p < 0.001$)
Social Space	-0.218*** [-0.303, -0.134] ($p < 0.001$)	-0.277*** [-0.369, -0.184] ($p < 0.001$)	-0.248*** [-0.339, -0.158] ($p < 0.001$)	-0.245*** [-0.336, -0.155] ($p < 0.001$)	-0.277*** [-0.374, -0.180] ($p < 0.001$)
Subfield Growth					
Cum. Subfield Size (logged)		0.240*** [0.168, 0.313] ($p < 0.001$)	0.726*** [0.602, 0.851] ($p < 0.001$)	0.739*** [0.613, 0.864] ($p < 0.001$)	0.499*** [0.334, 0.664] ($p < 0.001$)
2-year Subfield Growth		-0.116*** [-0.130, -0.103] ($p < 0.001$)	-0.197*** [-0.214, -0.179] ($p < 0.001$)	-0.197*** [-0.215, -0.180] ($p < 0.001$)	-0.217*** [-0.237, -0.198] ($p < 0.001$)
Citation Dynamics					
Cum. Citations (logged)			-2.437*** [-2.675, -2.199] ($p < 0.001$)	-2.458*** [-2.700, -2.215] ($p < 0.001$)	-2.472*** [-2.794, -2.149] ($p < 0.001$)
2-year Citations (logged)			2.378*** [2.128, 2.629] ($p < 0.001$)	2.386*** [2.134, 2.638] ($p < 0.001$)	2.772*** [2.485, 3.058] ($p < 0.001$)
Gini Coef. of Cum. Citation			0.011 [0.000, 0.022] ($p = 0.050$)	0.012* [0.001, 0.023] ($p = 0.04$)	0.012 [-0.001, 0.024] ($p = 0.065$)
Gini Coef. of 2-year Citation			-0.018*** [-0.028, -0.009] ($p < 0.001$)	-0.018*** [-0.028, -0.009] ($p < 0.001$)	-0.026*** [-0.037, -0.015] ($p < 0.001$)
Other Controls					
Retraction Notice Published				-0.035 [-0.405, 0.335] ($p = 0.853$)	0.062 [-0.328, 0.453] ($p = 0.754$)
After Death				0.288 [0.012, 0.563] ($p = 0.041$)	0.152 [-0.100, 0.404] ($p = 0.236$)
After Death * Superstar Death				-0.186** [-0.322, -0.05] ($p = 0.007$)	-0.138 [-0.306, 0.031] ($p = 0.109$)
Fixed Effects					
Calendar Year	N	N	N	N	Y
Strata ID	N	N	N	N	Y
Log-Likelihood	-31,686.1	-30,832.3	-28,989.6	-28,978.4	-26,289.5
Total Observations	1,313,433				

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table A2.2: Model Estimates with the Bottom 0.25% Cutoff, with and without Controls

Dependent Variable	*Substantial Decline of Citation*				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Knowledge Diffusion					
Scientific Space	-0.234*** [-0.323, -0.145] ($p < 0.001$)	-0.289*** [-0.382, -0.196] ($p < 0.001$)	-0.140*** [-0.220, -0.061] ($p < 0.001$)	-0.141*** [-0.221, -0.061] ($p < 0.001$)	-0.221*** [-0.302, -0.139] ($p < 0.001$)
Social Space	-0.260*** [-0.363, -0.156] ($p < 0.001$)	-0.323*** [-0.438, -0.209] ($p < 0.001$)	-0.289*** [-0.400, -0.178] ($p < 0.001$)	-0.285*** [-0.396, -0.174] ($p < 0.001$)	-0.313*** [-0.431, -0.196] ($p < 0.001$)
Subfield Growth					
Cum. Subfield Size (logged)		0.276*** [0.190, 0.361] ($p < 0.001$)	0.830*** [0.681, 0.979] ($p < 0.001$)	0.846*** [0.697, 0.995] ($p < 0.001$)	0.649*** [0.408, 0.890] ($p < 0.001$)
2-year Subfield Growth		-0.126*** [-0.145, -0.107] ($p < 0.001$)	-0.211*** [-0.234, -0.189] ($p < 0.001$)	-0.212*** [-0.234, -0.19] ($p < 0.001$)	-0.240*** [-0.266, -0.214] ($p < 0.001$)
Citation Dynamics					
Cum. Citations (logged)			-2.624*** [-2.900, -2.347] ($p < 0.001$)	-2.648*** [-2.927, -2.370] ($p < 0.001$)	-2.739*** [-3.098, -2.380] ($p < 0.001$)
2-year Citations (logged)			2.545*** [2.253, 2.836] ($p < 0.001$)	2.555*** [2.263, 2.847] ($p < 0.001$)	2.976*** [2.652, 3.301] ($p < 0.001$)
Gini Coef. of Cum. Citation			0.011 [-0.005, 0.028] ($p = 0.172$)	0.012 [-0.004, 0.029] ($p = 0.153$)	0.011 [-0.007, 0.029] ($p = 0.241$)
Gini Coef. of 2-year Citation			-0.021** [-0.035, -0.008] ($p = 0.002$)	-0.021** [-0.035, -0.008] ($p = 0.002$)	-0.029*** [-0.044, -0.013] ($p < 0.001$)
Other Controls					
Retraction Notice Published				-0.166 [-0.653, 0.321] ($p = 0.503$)	-0.069 [-0.594, 0.456] ($p = 0.796$)
After Death				0.404* [0.026, 0.782] ($p = 0.036$)	0.263 [-0.078, 0.604] ($p = 0.130$)
After Death * Superstar Death				-0.205* [-0.380, -0.031] ($p = 0.021$)	-0.150 [-0.371, 0.071] ($p = 0.182$)
Fixed Effects					
Calendar Year	N	N	N	N	Y
Strata ID	N	N	N	N	Y
Log-Likelihood	-18,129.3	-17,635.5	-16,532.5	-16,523.1	-14,527.7
Total Observations	1,366,970				

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table A2.3: Model Estimates with the Bottom 0.1% Cutoff, with and without Controls

Dependent Variable	*Substantial Decline of Citation*				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Knowledge Diffusion					
Scientific Space	-0.209** [-0.342, -0.076] (p = 0.002)	-0.263*** [-0.404, -0.122] (p < 0.001)	-0.105 [-0.228, 0.017] (p = 0.091)	-0.107 [-0.23, 0.016] (p = 0.090)	-0.179** [-0.297, -0.060] (p = 0.003)
Social Space	-0.344*** [-0.506, -0.183] (p < 0.001)	-0.419*** [-0.603, -0.235] (p < 0.001)	-0.376*** [-0.550, -0.203] (p < 0.001)	-0.371*** [-0.544, -0.197] (p < 0.001)	-0.435*** [-0.642, -0.229] (p < 0.001)
Subfield Growth					
Cum. Subfield Size (logged)		0.323*** [0.197, 0.448] (p < 0.001)	0.967*** [0.791, 1.142] (p < 0.001)	0.992*** [0.814, 1.169] (p < 0.001)	0.980*** [0.683, 1.278] (p < 0.001)
2-year Subfield Growth		-0.139*** [-0.171, -0.107] (p < 0.001)	-0.227*** [-0.258, -0.196] (p < 0.001)	-0.228*** [-0.26, -0.197] (p < 0.001)	-0.271*** [-0.308, -0.234] (p < 0.001)
Citation Dynamics					
Cum. Citations (logged)			-2.751*** [-3.100, -2.401] (p < 0.001)	-2.793*** [-3.147, -2.44] (p < 0.001)	-3.117*** [-3.545, -2.689] (p < 0.001)
2-year Citations (logged)			2.612*** [2.232, 2.992] (p < 0.001)	2.626*** [2.246, 3.006] (p < 0.001)	3.111*** [2.684, 3.538] (p < 0.001)
Gini Coef. of Cum. Citation			0.011 [-0.022, 0.043] (p = 0.515)	0.012 [-0.021, 0.044] (p = 0.482)	0.017 [-0.015, 0.05] (p = 0.298)
Gini Coef. of 2-year Citation			-0.019 [-0.047, 0.009] (p = 0.178)	-0.019 [-0.047, 0.009] (p = 0.179)	-0.025 [-0.058, 0.008] (p = 0.131)
Other Controls					
Retraction Notice Published				0.224 [-0.469, 0.916] (p = 0.527)	0.385 [-0.377, 1.147] (p = 0.322)
After Death				0.764*** [0.318, 1.211] (p < 0.001)	0.556* [0.073, 1.039] (p = 0.024)
After Death * Superstar Death				-0.289 [-0.584, 0.005] (p = 0.054)	-0.163 [-0.574, 0.249] (p = 0.438)
Fixed Effects					
Calendar Year	N	N	N	N	Y
Strata ID	N	N	N	N	Y
Log-Likelihood	-8,107.5	-7,879.7	-7,405.3	-7,395.6	-6,010.4
Total Observations	1,401,037				

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Measuring Knowledge Diffusion Through Document Embedding Spaces

We train vector representation models for biomedical science publications from PKG 2020 to locate positions of scientific publications based on their contents and to measure the similarity/distance between papers linked through citations. We adapt the Doc2vec model (Le and Mikolov 2014), a variant of the Word2vec model (Mikolov et al. 2013; Garg et al. 2018), which was initially developed to produce dense vector representations for documents or paragraphs from the words that compose them. Word embedding models generate a high-dimensional vector space in which geometrically proximate word vectors correspond to words that frequently share local linguistic contexts in the training data (Mikolov et al. 2013; Garg et al. 2018; Kozlowski, Taddy, and Evans 2019). This approach has previously been extended to generate representational vectors for entities connected in networks by substituting connections among entities as shared contexts (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016).

We consider that a research article can be characterized by 1) a list of MeSH terms and 2) researchers authoring it. Accordingly, we build two separate representational vector spaces — "scientific space" and "social space." We employ the Python Gensim package (version 4.0) (Radim Rehurek 2010) to train our vector representations. We specifically use the Distributed Bag of Words (DBOW) model, analogous to the skip-gram model from the Word2vec framework to train document vectors and constituting elements (MeSH terms and author IDs) simultaneously. This approach enables us to conduct document retrieval tasks using the vector representations of MeSH terms and author IDs to validate the resulting spaces. Detailed implementation procedures are as follows.

Scientific Space from MeSH Descriptors

We posit the MeSH terms as constituting words to build a "scientific space" for the biomedical literature. Because nominal terminologies are subject to change, we use MeSH terms' unique IDs from the National Library of Medicine. For instance, a MeSH descriptor, *Mesenchymal Stem Cells* (Descriptor ID: *D059630*), was indexed as *Mesenchymal Stromal Cells* from 2012 to 2018. However, it began to be reindexed as *Mesenchymal stem cells* in 2019, while its uniquely assigned descriptor ID, *D059630*, remains the same.

Figure A2.1: MeSH terms assigned to PMID 28376884

When a MeSH qualifier is attached to a MeSH descriptor, we consider both a descriptor with a qualifier and without it. Note that Figure A2.1 displays MeSH terms assigned to "Cancer immunotherapies targeting the PD-1 signaling pathway" (PMID 28376884), published in the *Journal of Biomedical Science* in 2017, authored by Iwai, Hamanishi, Chamoto, and Honjo(Iwai et al. 2017). The second term, *Antineoplastic Agents / metabolism*, can be broken down into the primary MeSH descriptor, *Antineoplastic Agents*, and the qualifier, *metabolism*, narrowing down the scope. The third term, *Antineoplastic Agents / pharmacology**, also has a qualifier, *pharmacology*. (The asterisk denotes that the given term is a major topic of the publication.) For this case, we include 1) *Antineoplastic Agents,* 2) *Antineoplastic Agents / metabolism,* and 3) *Antineoplastic Agents / pharmacology* for our model training. We do this to reflect that PubMed search queries using only MeSH terms (without qualifiers), *Antineoplastic Agents*, for this case,

capture publications like PMID 28376884. We exclude the asterisks for the same reason, taking into consideration co-searchability. As a result, the final list of MeSH terms fed into the training process for PMID 28376884 is A*ntibodies, Monoclonal*, *Antibodies*, *Monoclonal / therapeutic use*, *Antineoplastic Agents*, *Antineoplastic Agents / metabolism*, *Antineoplastic Agents / pharmacology*, *Humans*, *Immunologic Factors*, *Immunologic Factors / therapeutic use*, *Immunotherapy*, *Neoplasms*, *Neoplasms / therapy*, *Programmed Cell Death 1 Receptor*, *Programmed Cell Death 1 Receptor / therapeutic use*, *Signal Transduction*.

With these MeSH combinations, we train 100-dimensional vectors for 26,666,615 PMIDs and 303,492 MeSH combinations that appear at least ten times with 100 training epochs. The mean number of MeSH terms (after the procedure detailed above) per PMID from our dataset is 16.34 (std=9.04). However, we set the sliding window size that defines the boundary of the training context as 110, the maximum number from the data, to ensure that each training instance includes all the other MeSH combinations on a given article without splitting them up by imposing arbitrary contexts.

We validate the resulting vector representations by attempting to retrieve resulting publication vectors using MeSH combination vectors across 20 random samples, each containing 1,000 publications. We first take the vectors of MeSH terms assigned to each publication, infer the position of a document combining the MeSH terms, and check its proximity to the original vector representation of the article containing those MeSH terms. It is, for instance, a test to see if we can retrieve PMID 28376884 in Figure A2.1 by inferring the position of a document combining the vectors of MeSH terms assigned to it. Because it is impossible to differentiate publications with the same set of MeSH terms with this model, we consider the 1, 5, and 10 most similar documents from the inferred vector, using cosine similarity. We find that it is possible to

retrieve the target PMIDs with the rate of 92.48% (SD= .81), 96.14% (SD=.59), and 97.18% (SD=.52) from the top 1, 5, and 10 most similar documents, respectively, which suggests documents sharing MeSH terms are located close together in the 100-dimensional embedding space.

An advantage of using this Doc2vec model is that it reflects the high-order proximity of constituting words beyond their direct co-occurrence in a context. Consider two documents, PMID 23142641, a review article titled "Challenges measuring cardiomyocyte renewal," published in 2013 (Soonpaa, Rubart, and Field 2013), and PMID 11287958, an original research article, "Bone marrow cells regenerate infarcted myocardium," published in 2001 (Orlic et al. 2001). The former review article cited the latter article. A simple but popular similarity metric would be the Jaccard coefficient ranging from 0 to 1, computed by dividing the number of MeSH terms that two articles share by the size of the union set of all MeSH terms assigned to the two publications.

The MeSH terms assigned to PMID 23142641 are *Animals; Bromodeoxyuridine; Cell Differentiation; Cell Nucleus / metabolism; Cell Nucleus / ultrastructure; Cell Proliferation; Cell Tracking; Genes, Reporter; Integrases; Mice; Mice, Transgenic; Myocardium / cytology*; Myocardium / metabolism; Myocytes, Cardiac / cytology*; Myocytes, Cardiac / metabolism; Regeneration; Stem Cells / cytology*; Stem Cells / metabolism; beta-Galactosidase*.

The MeSH terms assigned to PMID 11287958 are as follows: *Animals; Bone Marrow Transplantation*; Cell Differentiation; Connexin 43 / metabolism; DNA-Binding Proteins / metabolism; Female; Green Fluorescent Proteins; Ki-67 Antigen / metabolism; Luminescent Proteins / metabolism; MEF2 Transcription Factors; Male; Mice; Mice, Inbred C57BL; Mice, Transgenic; Myocardial Infarction / therapy*; Myocardium / cytology; Myocardium /*

pathology; Myogenic Regulatory Factors; Proto-Oncogene Proteins c-kit / metabolism; Transcription Factors / metabolism*.

The Jaccard coefficient of the two publications based on the MeSH terms is .133 despite the close relationship between the two articles. However, the cosine similarity between the two documents on our trained model is .844, which better reflects the overall topic similarity between the two publications.

Social Space with Disambiguated Author IDs

Analogous to the content embedding space from MeSH terms, we also build a 100-dimensional social embedding space using Doc2vec, anchored by 8,359,189 disambiguated biomedical authors, within which we locate the vector space position of 28,329,992 PMIDs published by the end of 2019. In other words, we consider the author IDs as constituting document units. To inscribe the co-author information per publication, we included only authors that appeared more than once. The mean number of authors per publication from 28,329,992 PMIDs is 3.97 (std=5.01) with a median of 3. However, we set the window size for the training context as 2000 – arbitrarily larger than the maximum number of authors in the dataset – to include all author IDs in the training process for a given publication. We do this to ensure that the resulting article embedding model assigns similar vectors to articles co-authored by the same groups of overlapping co-authors who are directly or indirectly close in the social space of biomedical research collaboration. We trained our social embedding space using 100 epochs (or training iterations).

We validate the quality of vector representations in the same manner we did for the MeSH content space across 20 random samples of 1,000 publications each. We take the author

vectors for each publication, infer the position of a hypothetical publication those authors could have written within the 100-dimensional embedding space, and check its proximity to the vector representation written by the same author(s). Considering the impossibility of distinguishing publications written by the same author(s), we also assess the 1, 5, 10, and 20 most similar PMIDs from the inferred vector using cosine similarity. The target PMIDs could be retrieved with the rate of 65.26% (SD= 1.73), 86.16% (SD=1.06), 90.27% (SD=0.74), 92.9% (SD=0.77) from the top 1, 5, 10, 20 most similar documents, respectively. The sharp increase in self-retrieval for relaxed conditions demonstrates that papers written by the same author(s) are contiguous in the resulting 100-dimensional social embedding space.

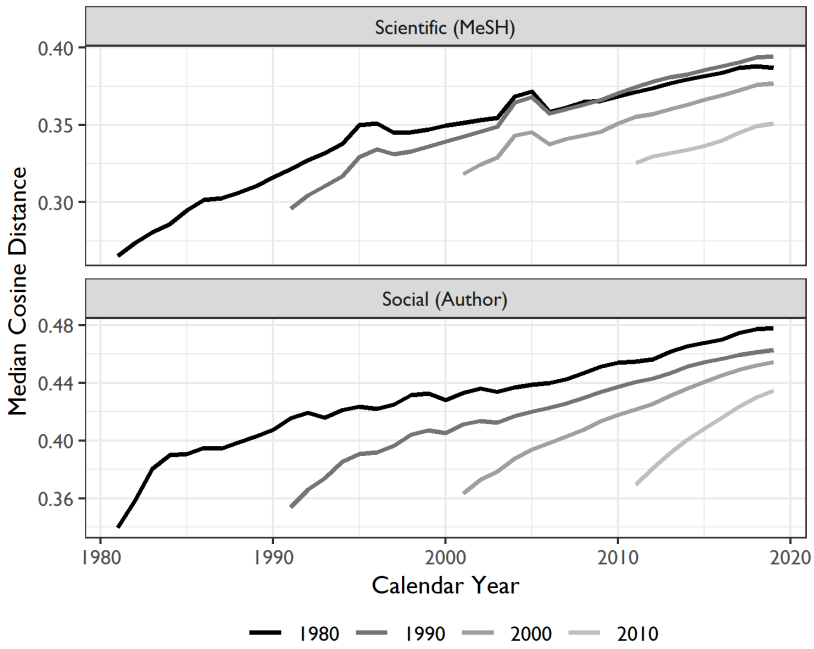
Aggregated Pattern for Diffusion for Highly Cited Articles Published in 1980, 1990, 2000, 2010

Here, we provide an aggregate-level description of how our diffusion indices temporally evolve using highly cited papers (top 5% percent in citation counts by the end of 2019) from four cohorts of research articles published in 1980, 1990, 2000, and 2010. We first make subsets of publications that the raw citation obtained by the end of 2019 fall over the 5% percentile in each cohort year (10,967 of 219,358 in 1970; 14,031 of 280,622 in 1980; 20,527 of 410,555 in 1990; 26,513 of 530,271 in 2000; 41,156 /823,129 in 2010), also accordingly extract cosine distances between the focal papers and citing papers measured in social and scientific space. With data from two rolling years, medians of cosine distances from two spaces each calendar year are computed. For example, the median cosine distance assigned to 1991 for the 1990 cohort is computed using all the citations observed in 1990 and 1991.

Figure A2.2 shows the temporal evolution of diffusion metrics from scientific and social space. As the universe of biomedical entities and scientists expands, distances between focal

papers and citing papers tend to increase in both scientific and social spaces by 2019. Then, the pattern, especially from the scientific space, indicates that our 100-dimensional representational spaces may allocate publications in some years (e.g., 2004 and 2005) in relatively distant locations within a trained manifold in the training process. Together, these suggest a necessity to consider the calendar year effect when a research article was published for the following analysis.

Figure A2.2: Temporal pattern of diffusion from highly cited articles (Top 5%) published in 1980, 1990, 2000, 2010



Note: Median cosine distances for each year (t) are computed based on a two-year rolling (t and t-1) window.

Delineating Biomedical Subfield

Science is a social enterprise: like any other intellectual product, biomedical science attains its significance when others recognize and build upon it (Bourdieu 1975; McMahan and McFarland 2021; Crane 1972). Hence, we seek to understand the dynamics of diffusion and shifting attention beyond individual publications at the subfield level. We utilize the 'Similar Article' (or 'Related Articles') function provided by PubMed (Lin and Wilbur 2007), powered

by the Pubmed Related Article algorithm (PMRA) (Lin and Wilbur 2007), which uses words in the abstracts, titles, and MeSH terms to capture a set of intellectually neighboring articles from a given seed article. This approach has been employed to study the repercussions of scientific scandals on careers (Azoulay, Bonatti, and Krieger 2017), shifts in research focus among scientists responding to NIH funding changes (Azoulay, Bonatti, and Krieger 2017; Myers 2020), and the negative impact of winning prizes for recipient competitors (Reschke, Azoulay, and Stuart 2018).

We apply this method to seed articles curated by Azoulay and their colleagues (2019), which consist of research papers from U.S. elite life scientists published between 1970 and 2002 (inclusive). The authors deemed scientist's elite if they satisfied one or more of the following criteria: they were (i) highly funded, (ii) highly cited, (iii) top patents, (iv) members of the National Academy of Sciences, (v) the National Academy of Medicine, or (vi) early career prize winners (i.e., NIH MERIT awardees, Howard Hughes Medical Investigators). To estimate the effects of the premature death of elite biomedical researchers, the study first identified 3,076 seed articles authored by 452 researchers who died prematurely. These 452 researchers represent a subset of a larger pool of 12,935 star scientists, with a median (and mean) age at death of 61. Within this group, 229 passed away following a protracted illness, while 185 died suddenly and unexpectedly (e.g., in a car crash). Forty percent of these stars held an M.D. rather than a Ph.D.; 90 percent were male, and each received an average of $16.6 million in NIH grants and published 138 papers, garnering 8,341 citations over their careers.

The study then performed 'coarsened exact matching' to identify a control group of publications from elite scientists who did not experience premature death, taking into account factors such as 1) publication years, 2) team sizes, 3) the ages of the elite scientists, and 4) long-

run citation impact. This process allows a subfield to be matched to several subfields associated with the early death of eminent scientists, leading to overlaps. Consequently, 4,180 subfields were matched to more than one subfield with the premature death of associated superstar scientists, leading to duplicates when counted individually. This results in 34,218 unique pairs of subfield strata and seed articles for subfield identification, for 28,504 unique seed articles spanning subfields. We repurpose this dataset to investigate a different question from the original study in our work.

Consistent with the original approach, we assume each seed article represents distinct subfields in biomedicine, but we extended the period to the end of 2019 as their subfield panel data stopped in 2006. We identify 1,941,680 unique publications (including the seed articles) spanning 28,504 unique subfields. By the end of 2019, the mean and median size of subfields is 122.52 and 102, respectively, with a standard deviation of 91.5. Table A2.4 below shows the subfield sizes at the 1st, 10th, 25th (1Q), 50th (median), 75th (3Q), 90th, and 99th percentiles at the end of 2019. Table A2.5 shows the distribution of citations garnered by 28,504 subfields by the end of 2019.

Table A2.4: Distribution of Subfield Size in 2019

Percentile	1st	10th	25th	50th	75th	90th	99th
Subfield Size in 2019	40	68	89	102	125	184	484

Table A2.5: Distribution of Cumulative Citations per Subfield at the End of 2019

Percentile	1st	10th	25th	50th	75th	90th	99th
Cumulative # of citations by the end of 2019	1,015	2,319	3,423	5,097	7,506	11,065	24,312

We extract research articles that have cited any 1,941,680 publications from the PKG 2020 citation database, which returns 11,421,194 publications and 86,804,637 paper-to-paper citations. Not all publications are associated with MeSH or Author IDs from PKG. We identify

10,894,779 publications that PKG assigns author IDs, constituting 84,389,548 citations from social space and 10,454,104 publications associated with MeSH terms linked through 82,228,828 citations.

Alternative Identification of Subfields

While PMRA underpins the PubMed interface, serving as a crucial tool for researchers to locate information related to focal research papers and study the operation of biomedical science(Azoulay, Bonatti, and Krieger 2017; Azoulay, Fons-Rosen, and Zivin 2019; Myers 2020; Reschke, Azoulay, and Stuart 2018), our analysis relies on this specific method of subfield identification. To ensure the robustness of our results, we have undertaken the following steps: 1) Using the same set of 28,504 seed article PMIDs and the scientific embedding space trained on MeSH terms, we redefined subfields by selecting the top *N* most similar articles to the seed articles based on cosine similarity within the scientific space, where '*N*' corresponds to the original subfield size as determined by PMRA. For the 82 out of 28,504 seed articles without assigned MeSH terms, we substituted each with a PMRA-identified similar article that ranked in the top 10 in similarity and had the smallest difference in publication years; 2) As an additional robustness check, we doubled the size of each subfield to assess the sensitivity of our results to changes in subfield sizes identified by PMRA.

With these two alternatively defined systems of subfields, we recalculated subfield-level variables consistent with our original analysis. We applied the same approach using three thresholds (i.e., 0.5%, 0.25%, 0.1%) to identify the sudden declines from two alternative subfields of the same and doubled sizes defined within our scientific embedding space. The following Tables report a pattern of results similar to the main findings.

Table A2.6: Estimates with Alternative Subfields using the Same Size of the Originals

Dependent Variable	***Substantial Decline of Citation***		
	Model 1 (0.5%)	**Model 2 (0.25%)**	**Model 3 (0.1%)**
Knowledge Diffusion			
Scientific Space	-0.250*** [-0.312, -0.188] ($p < 0.001$)	-0.289*** [-0.364, -0.213] ($p < 0.001$)	-0.358*** [-0.445, -0.270] ($p < 0.001$)
Social Space	-0.098** [-0.167, -0.03] ($p = 0.005$)	-0.113* [-0.205, -0.02] ($p = 0.017$)	-0.150* [-0.293, -0.006] ($p = 0.041$)
Log-Likelihood	-33,501.0	-19,308.5	-8,780.9
Total Observations	1,324,948	1,385,980	1,425,731

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table A2.7: Estimates with Controls for Alternative Subfields using the Same Size

Dependent Variable	***Substantial Decline of Citation***		
	Model 1 (0.5%)	**Model 2 (0.25%)**	**Model 3 (0.1%)**
Knowledge Diffusion			
Scientific Space	-0.258*** [-0.321, -0.194] ($p < 0.001$)	-0.276*** [-0.356, -0.196] ($p < 0.001$)	-0.339*** [-0.471, -0.208] ($p < 0.001$)
Social Space	-0.121** [-0.204, -0.038] ($p = 0.004$)	-0.157** [-0.263, -0.052] ($p = 0.003$)	-0.201* [-0.360, -0.043] ($p = 0.013$)
Subfield Growth			
Cum. Subfield Size (logged)	0.686*** [0.539, 0.833] ($p < 0.001$)	0.781*** [0.561, 1.000] ($p < 0.001$)	0.716*** [0.388, 1.043] ($p < 0.001$)
2-year Subfield Growth	-0.237*** [-0.256, -0.217] ($p < 0.001$)	-0.244*** [-0.269, -0.218] ($p < 0.001$)	-0.258*** [-0.291, -0.225] ($p < 0.001$)
Citation Dynamics			
Cum. Citations (logged)	-2.113*** [-2.426, -1.800] ($p < 0.001$)	-2.291*** [-2.648, -1.933] ($p < 0.001$)	-2.443*** [-2.863, -2.024] ($p < 0.001$)
2-year Citations (logged)	2.152*** [1.881, 2.422] ($p < 0.001$)	2.278*** [1.960, 2.595] ($p < 0.001$)	2.440*** [2.042, 2.838] ($p < 0.001$)
Gini Coef. of Cum. Citation	0.262 [-0.986, 1.510] ($p = 0.680$)	0.341 [-1.466, 2.148] ($p = 0.711$)	-0.261 [-3.222, 2.700] ($p = 0.863$)
Gini Coef. of 2-year Citation	-0.598 [-1.688, 0.492] ($p = 0.282$)	-0.748 [-2.457, 0.961] ($p = 0.391$)	-1.046 [-3.976, 1.883] ($p = 0.484$)

Table A2.7 continued

Other Controls			
Retraction Notice Published	-0.494 [-1.062, 0.074] (p = 0.088)	-0.262 [-1.22, 0.696] (p = 0.592)	0.299 [-0.873, 1.471] (p = 0.617)
After Death	0.191* [0.034, 0.348] (p = 0.017)	0.420*** [0.221, 0.619] (p < 0.001)	0.413 [-0.005, 0.831] (p = 0.053)
After Death * Superstar Death	0.017 [-0.143, 0.177] (p = 0.836)	0.027 [-0.169, 0.222] (p = 0.788)	-0.094 [-0.428, 0.240] (p = 0.581)
Fixed Effects			
Calendar Year	Y	Y	Y
Strata ID	Y	Y	Y
Log-Likelihood	-27,714.8	-15,542.6	-6,594.8
Total Observations	1,324,948	1,385,980	1,425,731

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * p < .05; ** p < .01; *** p < .001 (two-tailed)

Table A2.8: Estimates with Subfields from Scientific Space with Doubled Size of the Originals

Dependent Variable	***Substantial Decline of Citation***		
	Model 1 (0.5%)	**Model 2 (0.25%)**	**Model 3 (0.1%)**
Knowledge Diffusion			
Scientific Space	-0.268*** [-0.359, -0.177] (p < 0.001)	-0.344*** [-0.442, -0.247] (p < 0.001)	-0.395*** [-0.536, -0.255] (p < 0.001)
Social Space	-0.144*** [-0.215, -0.073] (p < 0.001)	-0.141** [-0.247, -0.034] (p = 0.009)	-0.196* [-0.360, -0.031] (p = 0.02)
Log-Likelihood	-34,063.1	-19,601.3	-8,827.1
Total Observations	1,389,103	1,446,414	1,484,148

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * p < .05; ** p < .01; *** p < .001 (two-tailed)

Table A2.9: Estimates with Controls for Alternative Subfields using the Double Size

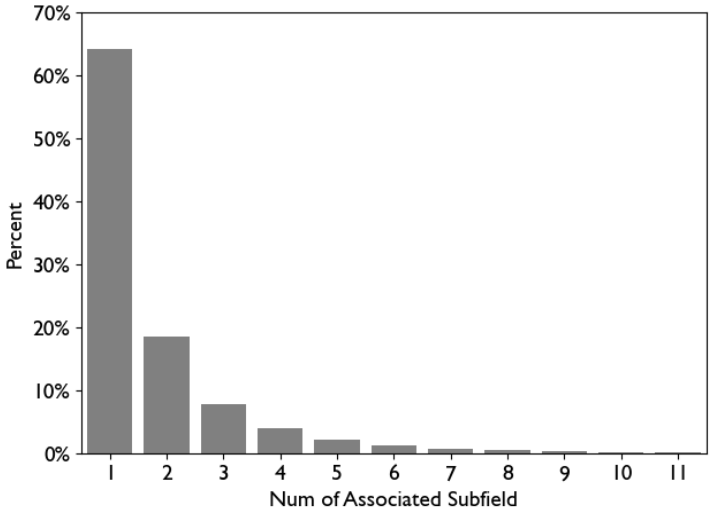
Dependent Variable	***Substantial Decline of Citation***		
	Model 1 (0.5%)	**Model 2 (0.25%)**	**Model 3 (0.1%)**
Knowledge Diffusion			
Scientific Space	-0.288*** [-0.364, -0.212] ($p < 0.001$)	-0.345*** [-0.432, -0.258] ($p < 0.001$)	-0.355*** [-0.523, -0.186] ($p < 0.001$)
Social Space	-0.178*** [-0.258, -0.097] ($p < 0.001$)	-0.208*** [-0.323, -0.093] ($p < 0.001$)	-0.265** [-0.439, -0.091] ($p = 0.003$)
Subfield Growth			
Cum. Subfield Size (logged)	0.603*** [0.422, 0.784] ($p < 0.001$)	0.679*** [0.447, 0.91] ($p < 0.001$)	0.883*** [0.487, 1.279] ($p < 0.001$)
2-year Subfield Growth	-0.307*** [-0.332, -0.282] ($p < 0.001$)	-0.320*** [-0.347, -0.293] ($p < 0.001$)	-0.352*** [-0.401, -0.304] ($p < 0.001$)
Citation Dynamics			
Cum. Citations (logged)	-2.542*** [-2.880, -2.204] ($p < 0.001$)	-2.738*** [-3.126, -2.349] ($p < 0.001$)	-3.204*** [-3.583, -2.825] ($p < 0.001$)
2-year Citations (logged)	2.532*** [2.241, 2.824] ($p < 0.001$)	2.678*** [2.334, 3.022] ($p < 0.001$)	3.029*** [2.649, 3.409] ($p < 0.001$)
Gini Coef. of Cum. Citation	1.989** [0.671, 3.306] ($p = 0.003$)	2.409* [0.52, 4.298] ($p = 0.012$)	2.447 [-1.217, 6.111] ($p = 0.191$)
Gini Coef. of 2-year Citation	-1.634* [-2.951, -0.317] ($p = 0.015$)	-1.896 [-3.932, 0.141] ($p = 0.068$)	-2.831 [-6.915, 1.252] ($p = 0.174$)
Other Controls			
Retraction Notice Published	-0.193 [-0.606, 0.221] ($p = 0.362$)	-0.102 [-0.76, 0.556] ($p = 0.762$)	0.290 [-0.776, 1.356] ($p = 0.594$)
After Death	0.200 [0.006, 0.393] ($p = 0.043$)	0.188 [-0.101, 0.477] ($p = 0.202$)	0.306 [-0.185, 0.797] ($p = 0.222$)
After Death * Superstar Death	-0.104 [-0.252, 0.043] ($p = 0.166$)	-0.115 [-0.316, 0.087] ($p = 0.264$)	-0.152 [-0.45, 0.145] ($p = 0.315$)
Fixed Effects			
Calendar Year	Y	Y	Y
Strata ID	Y	Y	Y
Log-Likelihood	-27,851.1	-15,571.3	-6,521.8
Total Observations	1,389,103	1,446,414	1,484,148

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Robustness Check with Mutually Exclusive Subfields

The PMRA-based subfield identification method allows a paper to be included in more than one subfield. The Azoulay team explored the extent of shared articles between pairs of PMRA-delineated subfields, focusing on 21,661 subfield pairs where a deceased superstar was last author on both associated source articles (see Figure C6 at aeaweb.org/content/file?id=10303). We conducted our investigation, noting that 1) the Azoulay team's analysis focused only on subfields associated with scientists who died prematurely, and 2) we extended the analysis window up to 2019. Our analysis shows that 64.1% of papers are predominantly associated with a single subfield.

Figure A2.3: Proportion of Papers by Number of Subfield Associated



We conducted additional analyses by reassigning overlapping papers to a single seed article, ensuring each subfield is mutually exclusive. We did this by leveraging distances within our scientific space. Specifically, for papers associated with more than one seed article, we compute cosine similarities between each paper and its seed articles and select the seed article with the highest similarity. After this procedure, we recomputed all subfield-level variables, including diffusion metrics, outcomes, and controls, in the same manner as in our original analyses. Results confirmed the robustness of our findings, as detailed below.

Table A2.10: Model Estimates after Reassignment Using the Bottom 0.50% Cutoff

Dep. Var.	*Substantial Decline of Citation*					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Know. Diff.						
Scientific Space	-0.204*** [-0.267, -0.140] ($p < 0.001$)	-0.232*** [-0.293, -0.170] ($p < 0.001$)	-0.128*** [-0.171, -0.086] ($p < 0.001$)	-0.129*** [-0.171, -0.088] ($p < 0.001$)	-0.182*** [-0.233, -0.131] ($p < 0.001$)	-0.218*** [-0.277, -0.160] ($p < 0.001$)
Social Space	-0.131*** [-0.185, -0.077] ($p < 0.001$)	-0.189*** [-0.249, -0.129] ($p < 0.001$)	-0.150*** [-0.211, -0.089] ($p < 0.001$)	-0.148*** [-0.208, -0.088] ($p < 0.001$)	-0.163*** [-0.229, -0.096] ($p < 0.001$)	-0.161*** [-0.227, -0.095] ($p < 0.001$)
Subfield Growth						
Cum. Subfield Size (logged)		0.309*** [0.249, 0.369] ($p < 0.001$	0.563*** [0.466, 0.661] ($p < 0.001$)	0.572*** [0.471, 0.673] ($p < 0.001$	0.151* [0.014, 0.287] ($p = 0.031$)	0.049 [-0.071, 0.170] ($p = 0.421$)
2-year Subfield Growth		-0.094*** [-0.104, -0.084] ($p < 0.001$	-0.143*** [-0.157, -0.129] ($p < 0.001$)	-0.143*** [-0.157, -0.130] ($p < 0.001$	-0.136*** [-0.150, -0.122] ($p < 0.001$)	-0.142*** [-0.158, -0.127] ($p < 0.001$)
Citation Dynamics						
Cum. Citations (logged)			-1.895*** [-2.154, -1.637] ($p < 0.001$)	-1.906*** [-2.168, -1.645] ($p < 0.001$)	-1.650*** [-1.955, -1.345] ($p < 0.001$)	-1.673*** [-1.957, -1.389] ($p < 0.001$)
2-year Citations (logged)			1.909*** [1.655, 2.163] ($p < 0.001$)	1.914*** [1.658, 2.169] ($p < 0.001$)	2.210*** [1.941, 2.479] ($p < 0.001$)	2.426*** [2.154, 2.698] ($p < 0.001$)
Gini Coef. of Cum. Citation			0.432 [-0.225, 1.089] ($p = 0.198$)	0.461 [-0.196, 1.119] ($p = 0.169$)	0.025 [-0.741, 0.791] ($p = 0.949$)	0.180 [-0.597, 0.957] ($p = 0.65$)
Gini Coef. of 2-year Citation			-0.511 [-1.131, 0.110] ($p = 0.107$)	-0.516 [-1.139, 0.108] ($p = 0.105$)	-1.420*** [-2.096, -0.745] ($p < 0.001$)	-1.593*** [-2.230, -0.956] ($p < 0.001$)
Other Controls						
Retraction Notice Published			-0.237 [-0.649, 0.175] ($p = 0.26$)	-0.077 [-0.531, 0.376] ($p = 0.738$)	0.01 [-0.54, 0.56] ($p = 0.972$)	-0.237 [-0.649, 0.175] ($p = 0.26$)
After Death			0.080 [-0.121, 0.28] ($p = 0.437$)	0.121 [-0.031, 0.274] ($p = 0.118$)	0.200* [0.032, 0.369] ($p = 0.02$)	0.08 [-0.121, 0.28] ($p = 0.437$)
After Death * Superstar Death			-0.149* [-0.283, -0.015] ($p = 0.029$	-0.002 [-0.168, 0.165] ($p = 0.984$)	-0.510** [-0.895, -0.125] ($p = 0.009$)	-0.149* [-0.283, -0.015] ($p = 0.029$)
Fixed Effects						
Calendar Year	N	N	N	N	Y	Y
Strata ID	N	N	N	N	Y	N
Star ID	N	N	N	N	N	Y
Log-Likelihood	-30,045.0	-29,187.9	-27,699.2	-27,694.0	-25,118.1	-22,709.7
Total Obs.	1,252,242					

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table A2.11: Model Estimates after Reassignment Using the Bottom 0.25% Cutoff

Dep. Var.	*Substantial Decline of Citation*					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Know. Diff.						
Scientific Space	-0.195*** [-0.273, -0.116] ($p < 0.001$)	-0.217*** [-0.295, -0.14] ($p < 0.001$)	-0.102*** [-0.159, -0.045] ($p < 0.001$)	-0.103*** [-0.159, -0.046] ($p < 0.001$)	-0.142*** [-0.211, -0.074] ($p < 0.001$)	-0.171*** [-0.244, -0.098] ($p < 0.001$)
Social Space	-0.168*** [-0.231, -0.105] ($p < 0.001$)	-0.236*** [-0.308, -0.164] ($p < 0.001$)	-0.195*** [-0.268, -0.121] ($p < 0.001$)	-0.193*** [-0.266, -0.12] ($p < 0.001$)	-0.196*** [-0.276, -0.117] ($p < 0.001$)	-0.193*** [-0.272, -0.114] ($p < 0.001$)
Subfield Growth						
Cum. Subfield Size (logged)		0.357*** [0.288, 0.427] ($p < 0.001$)	0.696*** [0.581, 0.811] ($p < 0.001$)	0.701*** [0.583, 0.819] ($p < 0.001$)	0.289*** [0.116, 0.461] ($p = 0.001$)	0.155 [-0.023, 0.333] ($p = 0.089$)
2-year Subfield Growth		-0.097*** [-0.109, -0.086] ($p < 0.001$)	-0.15*** [-0.166, -0.135] ($p < 0.001$)	-0.151*** [-0.166, -0.136] ($p < 0.001$)	-0.144*** [-0.16, -0.129] ($p < 0.001$)	-0.151*** [-0.167, -0.135] ($p < 0.001$)
Citation Dynamics						
Cum. Citations (logged)			-2.090*** [-2.347, -1.833] ($p < 0.001$)	-2.097*** [-2.353, -1.84] ($p < 0.001$)	-1.850*** [-2.153, -1.547] ($p < 0.001$)	-1.921*** [-2.241, -1.601] ($p < 0.001$)
2-year Citations (logged)			2.046*** [1.795, 2.297] ($p < 0.001$)	2.049*** [1.798, 2.300] ($p < 0.001$)	2.357*** [2.079, 2.635] ($p < 0.001$)	2.658*** [2.360, 2.956] ($p < 0.001$)
Gini Coef. of Cum. Citation			0.665 [-0.110, 1.439] ($p = 0.092$)	0.681 [-0.090, 1.452] ($p = 0.083$)	0.082 [-0.790, 0.955] ($p = 0.854$)	0.503 [-0.419, 1.424] ($p = 0.285$)
Gini Coef. of 2-year Citation			-0.759* [-1.48, -0.038] ($p = 0.039$)	-0.763* [-1.485, -0.04] ($p = 0.039$)	-1.650*** [-2.449, -0.852] ($p < 0.001$)	-2.077*** [-2.848, -1.305] ($p < 0.001$)
Other Controls						
Retraction Notice Published			-0.193 [-0.725, 0.339] ($p = 0.476$)	0.079 [-0.451, 0.609] ($p = 0.771$)	-0.007 [-0.7, 0.685] ($p = 0.983$)	-0.193 [-0.725, 0.339] ($p = 0.476$)
After Death			0.050 [-0.218, 0.318] ($p = 0.713$)	0.049 [-0.224, 0.321] ($p = 0.727$)	0.064 [-0.200, 0.328] ($p = 0.634$)	0.050 [-0.218, 0.318] ($p = 0.713$)
After Death * Superstar Death			-0.090 [-0.294, 0.115] ($p = 0.391$)	0.115 [-0.129, 0.359] ($p = 0.356$)	-0.151 [-0.652, 0.349] ($p = 0.554$)	-0.09 [-0.294, 0.115] ($p = 0.391$)
Fixed Effects						
Calendar Year	N	N	N	N	Y	Y
Strata ID	N	N	N	N	Y	N
Star ID	N	N	N	N	N	Y
Log-Likelihood	-16,969.8	-16,495.6	-15,626.5	-15,625.3	-13,682.5	-12,015.4
Total Obs.	1,304,469					

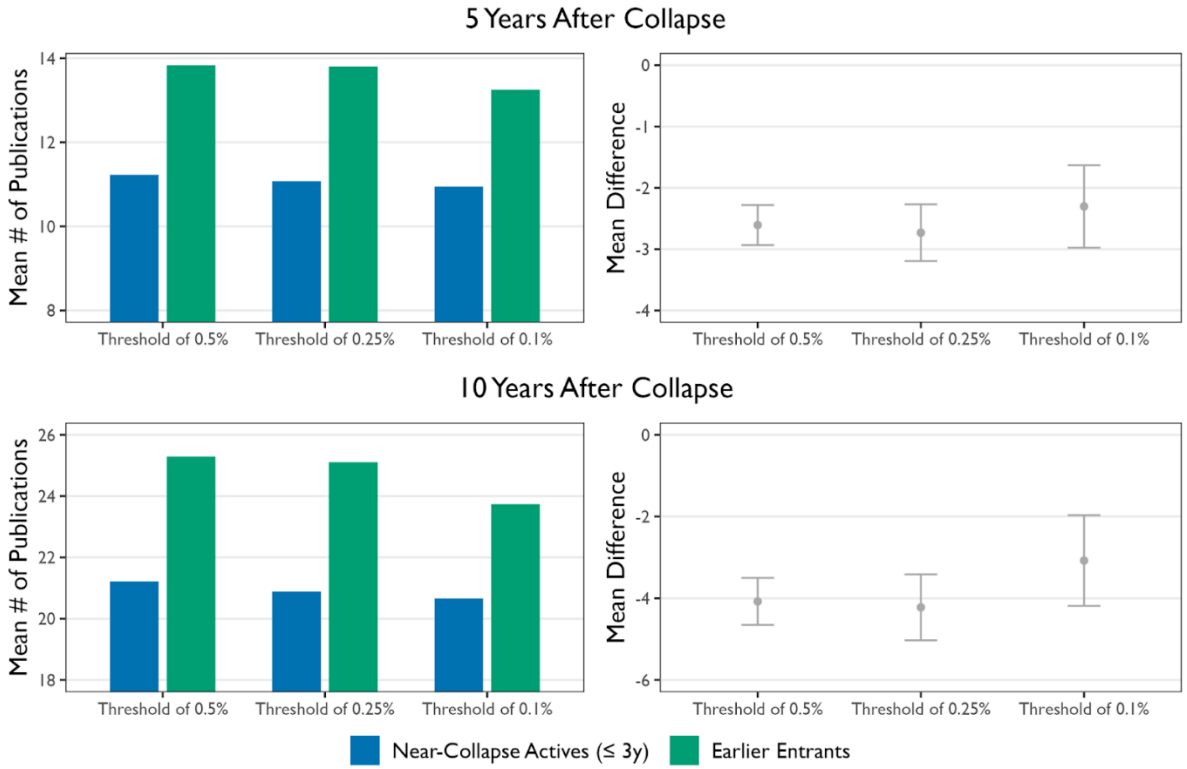
Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Table A2.12: Model Estimates after Reassignment Using the Bottom 0.10% Cutoff

Dep. Var.	*Substantial Decline of Citation*					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Know. Diff.						
Scientific Space	-0.243*** [-0.373, -0.113] ($p < 0.001$	-0.265*** [-0.400, -0.129] ($p < 0.001$)	-0.139* [-0.251, -0.028] ($p = 0.014$)	-0.140* [-0.251, -0.028] ($p = 0.014$)	-0.161* [-0.310, -0.013] ($p = 0.033$)	-0.229*** [-0.367, -0.091] ($p = 0.001$)
Social Space	-0.198*** [-0.298, -0.098] ($p < 0.001$	-0.271*** [-0.384, -0.158] ($p < 0.001$)	-0.229*** [-0.344, -0.114] ($p < 0.001$)	-0.226*** [-0.341, -0.111] ($p < 0.001$)	-0.238*** [-0.368, -0.109] ($p < 0.001$)	-0.186* [-0.331, -0.040] ($p = 0.012$)
Subfield Growth						
Cum. Subfield Size (logged)		0.385*** [0.294, 0.475] ($p < 0.001$)	0.778*** [0.613, 0.943] ($p < 0.001$)	0.786*** [0.618, 0.954] ($p < 0.001$)	0.326* [0.052, 0.601] ($p = 0.020$)	0.225 [-0.099, 0.55] ($p = 0.173$)
2-year Subfield Growth		-0.097*** [-0.114, -0.079] ($p < 0.001$)	-0.151*** [-0.170, -0.132] ($p < 0.001$)	-0.151*** [-0.171, -0.132] ($p < 0.001$)	-0.146*** [-0.167, -0.125] ($p < 0.001$)	-0.154*** [-0.175, -0.133] ($p < 0.001$)
Citation Dynamics						
Cum. Citations (logged)			-2.178*** [-2.514, -1.843] ($p < 0.001$)	-2.188*** [-2.525, -1.850] ($p < 0.001$)	-1.940*** [-2.350, -1.529] ($p < 0.001$)	-2.125*** [-2.594, -1.656] ($p < 0.001$)
2-year Citations (logged)			2.101*** [1.760, 2.443] ($p < 0.001$)	2.109*** [1.765, 2.453] ($p < 0.001$)	2.435*** [2.068, 2.803] ($p < 0.001$)	2.849*** [2.444, 3.253] ($p < 0.001$)
Gini Coef. of Cum. Citation			0.411 [-0.928, 1.751] ($p = 0.547$)	0.425 [-0.911, 1.762] ($p = 0.533$)	-0.053 [-1.657, 1.552] ($p = 0.949$)	0.594 [-0.958, 2.147] ($p = 0.453$)
Gini Coef. of 2-year Citation			-0.822 [-2.041, 0.396] ($p = 0.186$)	-0.838 [-2.061, 0.386] ($p = 0.180$)	-1.742** [-3.049, -0.435] ($p = 0.009$)	-2.515*** [-3.858, -1.172] ($p < 0.001$)
Other Controls						
Retraction Notice Published				-1.018 [-2.166, 0.13] ($p = 0.082$)	-0.614 [-1.912, 0.684] ($p = 0.354$)	-0.551 [-2.007, 0.905] ($p = 0.458$)
After Death				0.172 [-0.231, 0.575] ($p = 0.402$)	0.289 [-0.123, 0.701] ($p = 0.170$)	0.232 [-0.168, 0.633] ($p = 0.255$)
After Death * Superstar Death				-0.124 [-0.402, 0.154] ($p = 0.381$)	-0.024 [-0.395, 0.346] ($p = 0.898$)	0.297 [-0.948, 1.542] ($p = 0.64$)
Fixed Effects						
Calendar Year	N	N	N	N	Y	Y
Strata ID	N	N	N	N	Y	N
Star ID	N	N	N	N	N	Y
Log-Likelihood	-7,706.9	-7,511.0	-7,141.6	-7,138.5	-5,858.2	-4,901.2
Total Obs.	1,337,848					

Note: *Knowledge Diffusion* indices are standardized within field ages and calendar years across 28,504 subfields. The 95% confidence intervals inside brackets are computed based on the standard errors clustered at strata ID and calendar years. * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

Figure A2.4: Comparing Author Productivity in Collapsed Subfields 5- and 10-Years Post-Collapse after Reassignment



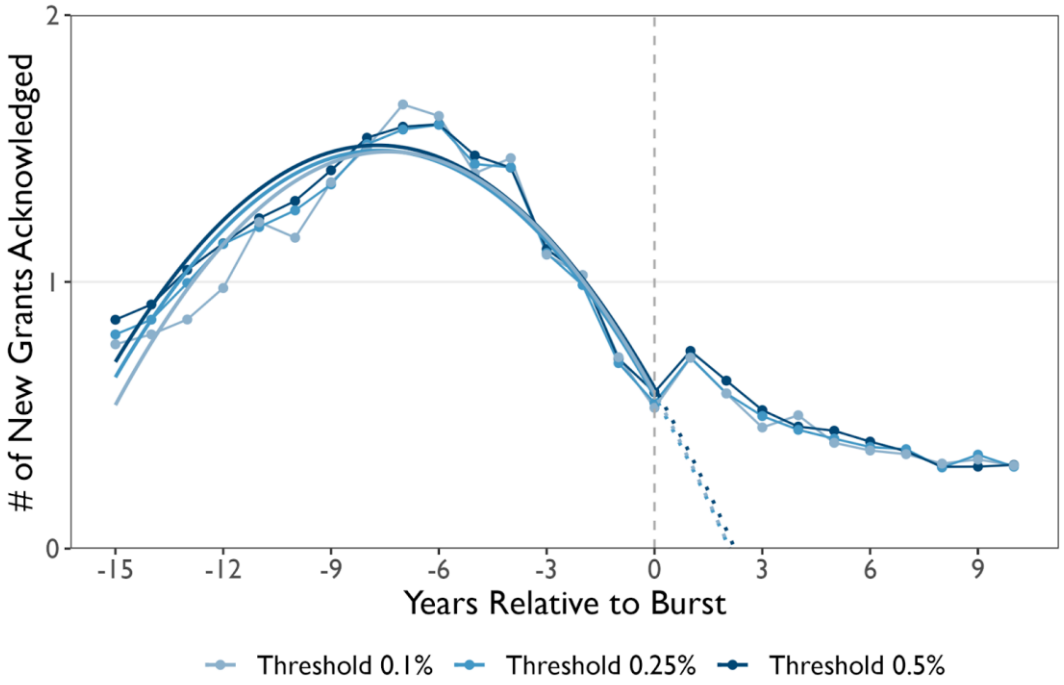
Note: Error bars represent 95% confidence intervals for the mean differences in average publication numbers. Comparisons are drawn between authors who entered the field early and those active near the collapse, based on paired *t*-tests.

Table A2.13: Pairwise *t*-test Comparing Average Productivity Differences between Near-Collapse Active Scientists (≤2 years before Collapse) and Early Entrants after Reassignment

	Threshold	Estimate	t	*p*-value (*d.f.*)	95% C.I.
5 Years	0.5%	-2.607	-15.60	< 0.001 (3,236)	[-2.280, -2.935]
	0.25%	-2.731	-11.57	< 0.001 (1,673)	[-2.269, -3.195]
	0.1%	-2.303	-6.72	< 0.001 (655)	[-1.631, -2.976]
10 Years	0.5%	-4.077	-13.87	< 0.001 (2,972)	[-3.500, -4.653]
	0.25%	-4.222	-10.25	< 0.001 (1,547)	[-3.415, -5.030]
	0.1%	-3.078	-5.45	< 0.001 (603)	[-1.968, -4.187]

Note: Subfields that collapsed after 2015 were excluded from the 5-year productivity comparison. Likewise, for the 10-year productivity, only subfields that collapsed on or before 2011 were included, considering the observation windows.

Figure A2.5: Average Number of New Grants Acknowledged in Collapsed Subfields by Years Relative to Burst after Reassignment



Note: The quadratic fit is applied to data from years -15 to 0 relative to the burst year, with dotted lines representing extrapolations starting from year 0 onwards.

Table A2.14: Proportion of Subfields with Newly Acknowledged Grants After Collapse, and the Mean, 1st Quantile, Median, and 3rd Quartile of the Number of New Grants Post-Collapse after Reassignment

Threshold	% of Subfields with New Grants Acknowledged After Collapse	Mean	Q1	Median	Q3
0.5%	68.45%	6.0	0	2	7
0.25%	68.79%	5.9	0	2	7
0.1%	68.68%	5.8	0	2	6

Note: Subfields that collapsed after 2015 were excluded from the 5-year productivity comparison. For the 10-year productivity analysis, only subfields that collapsed on or before 2011 were included, in consideration of the observation windows.

Chapter 3

Papers With Code or Without Code?

Impact of GitHub Repository Usability on the Diffusion of Machine Learning Research*

Abstract

Open Science initiatives prompt machine learning (ML) researchers and experts to share source codes - "scientific artifacts" - alongside research papers via public repositories such as GitHub. Here we analyze the extent to which 1) the availability of GitHub repositories influences paper citation and 2) the popularity trend of ML frameworks (e.g., PyTorch and TensorFlow) affects article citation rates. To accomplish this, we connect ML research publications indexed by Papers with Code (PwC) to Microsoft Academic Graph (MAG) and collect repository-level metadata using the GitHub API. Applying nearest-neighbor matching and econometric considerations, we estimate that papers enjoy approximately 20% advantages in monthly citation rates after the creation of the first GitHub repositories, accounting for paper-level fixed effects and ages. We also find that the temporal popularity trends for frameworks used in the first associated repositories could influence the monthly citation rate for papers. The results highlight the importance of technological artifacts and infrastructure underlying the diffusion of research.

---

* Co-authored by TaeYoung Kang and Junkyu Jang, the College of Business at the Korea Advanced Institute of Science and Technology. This chapter has been published in *Information Processing & Management* (Kang, Kang, and Jang 2023) and is reprinted with permission from Elsevier.

Introduction

In recent years, scientists and researchers have been encouraged, expected, or often required to share scientific artifacts along with their manuscripts—such as data and model implementations—via publicly accessible platforms. This desideratum is in line with the scientific movement (Frickel and Gross 2005; Peterson and Panofsky 2021) of *Open Science*, which has been fueled as a response to concerns regarding the reproducibility of various fields of research, ranging from medicine (Ioannidis 2005) to social sciences (Peterson and Panofsky 2021; Baker 2016). Recent commentaries have also suggested that the machine learning (ML) or, more broadly, artificial intelligence (AI) research community is not immune to these issues (Kapoor and Narayanan 2022; Hutson 2018; Haibe-Kains et al. 2020; Pineau et al. 2021). These concerns underscore the importance of research transparency, resonating with the ideal vision of modern science (Merton 1973).

While public sharing of research artifacts might not ultimately guarantee the reproducibility of a research product, consider a counterfactual where an ML research paper does not have an accompanying implementation. In such cases, results would not be readily testable, validated, and extended due to information asymmetry between knowledge initiators and the audience (Pavitt 1987). A multitude of procedures and complex computing environments are often latent in the codified descriptions in a research paper (Fonseca Cacho and Taghva 2018). As such, the lack of available code and repositories would diminish the value and relevance of ML research for other researchers and practitioners who build upon prior studies. Sharing scientific artifacts with research papers can yield social benefits by facilitating replication and validation (Mueller-Langer et al. 2019), and individual researchers can also increase the visibility

of research publications and garner more scientific credits by sharing associated scientific artifacts publicly (McKiernan et al. 2016).

The diffusion of research outputs and the allocation of academic credits, often assessed through the citation dynamics, represent a complex phenomenon influenced by numerous social and institutional factors (McMahan and McFarland 2021; Fortunato et al. 2018; Y. Huang et al. 2022; Min et al. 2021). Recent scholarship has investigated the incentives, motivations, and costs linked with research transparency (Kim and Adler 2015; Mukherjee and Stern 2009; Wilms et al. 2020), as well as the consequences of data and material sharing on the citation trajectories of research articles and associated implications (Kwon and Motohashi 2021; Furman and Stern 2011; Christensen et al. 2019). Building on this strand of literature, we aim to evaluate the degree to which the availability of code repositories, particularly GitHub repositories associated with papers, affects the citation rates of ML research articles by applying a combination of causal inference techniques (Dong et al. 2022).

Apart from emphasizing the importance of research transparency, we also aim to explore another critical aspect that could potentially influence the dissemination of ML research papers: the impact of ML framework choice for model development and implementation on citations of research papers. In recent years, frameworks such as PyTorch and TensorFlow have become essential tools for researchers and practitioners in the rapidly developing field of ML. Yet, to the best of our knowledge, their effect on research diffusion has rarely been systematically examined.

Motivated by the literature on the network effects (Katz and Shapiro 1986; Kauffman, McAndrews, and Wang 2000) and connecting the theory of cognitive shortcuts in decision-making (Tversky and Kahneman 1974) with the technology acceptance model (Davis 1989), we

probe whether the popularity of ML frameworks utilized in the first accompanying GitHub repositories for code implementation could affect the subsequent citation rates of ML research papers. We posit that the popularity of ML frameworks at the macro-level can generate a second-order network effect by invoking cognitive shortcuts for researchers under information overload, driven by the rapidly increasing volume of prior research. Extending this line of reasoning, we hypothesize that the popularity of a particular ML framework shapes the perceived ease of use and usefulness of research articles, thus boosting the citation rates of ML research papers whose models are implemented with more popular frameworks. To test this, we measure the monthly popularity of used frameworks, analogous to the "market shares" of these frameworks at the monthly interval within the code repositories cataloged by Papers with Code (PwC), and estimate the extent to which the popularity influences paper-level citations.

Our first analysis demonstrates a notable positive shift in citation rates after the first GitHub repositories associated with ML papers became publicly accessible. This finding underscores the importance of code repository availability in enhancing the visibility and impact of ML research papers. Subsequently, we reveal the effect of a framework's popularity on the paper citation stream. To do so, we link records from PwC to Microsoft Academic Graph (MAG) (Wang et al. 2019; Sinha et al. 2015), gather relevant repository-level information using GitHub API for approximately ~20K randomly selected papers, and apply the nearest-neighbor matching and two-ways fixed effects. Our estimation from the first analysis reveals that ML research papers, on average, enjoy about 20% advantages in monthly citation rates after the creation of the first GitHub repositories. The second analysis shows that the temporal popularity trends for frameworks used in the first associated repositories may influence the monthly citation rate of related ML papers.

By analyzing the impact of code repositories for ML research on citation rates, our work contributes to a deeper understanding of the techniques and methodologies employed by researchers. On the one hand, our findings provide empirical evidence and insights that further support *Open Science* practices, consistent with prior studies. On the other hand, our analysis also emphasizes the nuanced role of ML frameworks in research dissemination and recognition beyond the focus on benchmark datasets (Koch et al. 2021; Paullada et al. 2021).

Related Literature and Hypotheses

Open Science, Code Availability, and Research Impact

The ideal of modern science envisions the unreserved open distribution of novel discoveries and findings from scientific research (Merton 1973). However, actual scientific practices have not always hewed to this normative characterization. Numerous factors interplay, including rewards and recognition toward priority (Merton 1957), the tension between the norm of communality—dictating that research products should belong to the scientific community—and the desire to secure control of discoveries (Mitroff 1974) for warding off potential competitors (Latour and Woolgar 1979).

Despite the recent advancements of protocols and guidelines, it has been consistently highlighted that the current reward structure in science does not sufficiently incentivize replication endeavors (Heesen 2018). For instance, the review process may prioritize theoretical novelty and model performance over the quality of supplementary materials (e.g., data, code implementation, and documentation), which are essential for ensuring research replicability. However, many factors, including incomplete documentation in research papers, diverse computing environments, and deprecated or mismatched auxiliary tools and functions used in

model implementations, can remain tacit despite the written descriptions presented in ML research papers (Fonseca Cacho and Taghva 2018). This implies that a research paper, despite its prominence, is but one among many sharable scientific artifacts.

Empirical analyses have documented the impact of research data sharing. In the context of biological science, Furman & Stern (2011) demonstrated that transferring biomaterials used in research to the Biological Resource Center increased the visibility of publications, measured in article citation counts. Likewise, an analysis focusing on genomics reported substantial citation benefits from data sharing (Piwowar and Vision 2013). Studies examining other scientific domains, such as astrophysics (Dorch, Drachen, and Ellegaard 2015) and astronomy (Henneken and Accomazzi 2011), also suggest positive effects of transparent data sharing on citations. This is not limited to natural sciences. Christensen et al. (2019), employing changes in data-sharing policy among economics and political science journals, revealed that publications that adhered to the policy shift and shared research data accumulated more citations than those that did not.

In addition to the data-sharing practices, scholarly attention has increasingly been directed to the role of code repositories, particularly in the context of ML research. Code repositories, hosted on platforms like GitHub, enable researchers to share their model implementations from research articles, making them readily accessible and executable. These repositories, at the very least, signal researchers' confidence in the robustness and validity of their work, suggesting that the work merits consideration within the research community. More practically, the availability of well-documented code and implementation can substantially reduce the time and effort required by other researchers to reproduce, validate, or build upon existing research (Tennant et al. 2017). The accessible and executable implementation shared via code repositories can help clarify any ambiguities that might arise from the written descriptions

alone (Stodden et al. 2016). This increased clarity and accessibility, which aligns with the tenet of the open-source software (von Krogh et al. 2012), can render the research more convincing and appealing to other scientists and practitioners, thereby leading to a greater research impact.

Several studies have examined the relationship between the availability of code repositories and the citation rates for ML research papers published at specific ML and AI conferences. Vandewalle (2012) analyzed the impact of code availability on citation rates of articles published in IEEE Transactions on Image Processing (TIP) from 2004 to 2006. The findings revealed that approximately 10% of papers shared their code online, and these papers received significantly more citations. Vandewalle's follow-up study (2019) on TIP papers published in 2017 showed that the proportion of papers increased to 24%, and they received, on average, double the citations of papers without code implementation online. Bonneel et al. analyzed 374 conference proceedings circulated in 2014, 2016, and 2018 SIGGRAPH, an Association for Computing Machinery (ACM) computer graphics conference and found a significant correlation between code availability and citation counts. Similarly, an analysis from Bhattarai et al. (2022) on proceedings from eight computer science conferences also showed a significant correlation between code repository availability and paper-level citation.

Building on previous research, we also posit that the availability of code repositories, particularly repositories in GitHub—a platform initially designed to host open-source software and social coding (Peng 2019)—now also serving as the most popular venue for sharing research-related artifacts, positively impacts the citation trajectory of research papers (Hypothesis 1). However, our study aims to do more than show associations (Dong et al. 2022). We seek to evaluate the extent to which public code repositories impact the citation trajectories of ML research articles. To accomplish this, we employ causal inference techniques, including

nearest-neighbor matching and two-way fixed effects, allowing us to generate a more precise estimate of the influence of code repository availability on citation rates.

Network Effects of Technological Infrastructure

The evolving landscape of scholarly communication has introduced a variety of factors that can influence the diffusion—or citation rates—of research products. In addition to the impact of the availability of GitHub repositories, our study attempts to examine a more nuanced aspect: whether the popularity of ML frameworks, such as PyTorch and TensorFlow, deployed in implementation can enhance the citation rates of ML research papers. While developed and maintained by corporations such as Meta (previously Facebook) and Google, these open-source frameworks are utilized by researchers and practitioners to facilitate their model development. However, to our best knowledge, their impact on research diffusion has not been systematically investigated.

The notion of network effect offers a critical theoretical lens for conceiving the influence of ML frameworks on paper citation rates. The network effect refers to a phenomenon wherein the value or utility of a technological product or service transcends its inherent quality, affected by the size of its user base or its compatibility with complementary goods and services (Katz and Shapiro 1985). Such effects can occur in any technological domain requiring a certain level of training (Katz and Shapiro 1986) and even lead to standardization in extreme cases, as often exemplified by the QWERTY keyboard layout (Arthur 1989; David 1985). Informed by this perspective, we consider that ML frameworks are subject to network effects, where their value is determined by the size of their user base and the level of compatibility and community support. More importantly, as technological infrastructures, these ML frameworks also can induce

secondary or indirect network effects (Economides and Salop 1992), implying a potential spillover of the popularity from a particular technological component to other interconnected entities across a broader system. Following this line of reasoning, we expect that the popularity of ML frameworks can influence the diffusion of ML research, especially when authors share their implementation for their papers deploying ML frameworks.

We also deem that the rapidly growing volume of academic papers (Fortunato et al. 2018) can cause cognitive overload for individual researchers and scientists (Chu and Evans 2021). The vast amounts of information may lead to "technostress," or the pressure to continually stay up-to-date (Ragu-Nathan et al. 2008) with the fast-evolving technological environment. The theory of cognitive shortcuts, such as heuristics (Tversky and Kahneman 1974), suggests that researchers confronted with information overload may attempt to alleviate cognitive stress by applying strategies to simplify the process. The technology acceptance model (Davis 1989; Venkatesh et al. 2003) proposed from the literature in information management systems also supports this view by highlighting cognitive factors—perceived usefulness of technologies and ease of use. Empirical studies demonstrated the importance of cognitive and perceptual dimensions in various settings ranging from fostering trust for new technologies in the case of the national identity system (Li, Hess, and Valacich 2008) to the adoption of mobile internet services (H.-W. Kim, Chan, and Gupta 2007). In our study context, we extend this view that the popularity of ML frameworks deployed in code implementation can serve as a cue for researchers to filter related literature, irrespective of its actual significance and citability. With cognitive restriction, researchers who attempt to build upon previous work or situate their work in the web of literature may attempt to avoid additional burdens posed by research products that

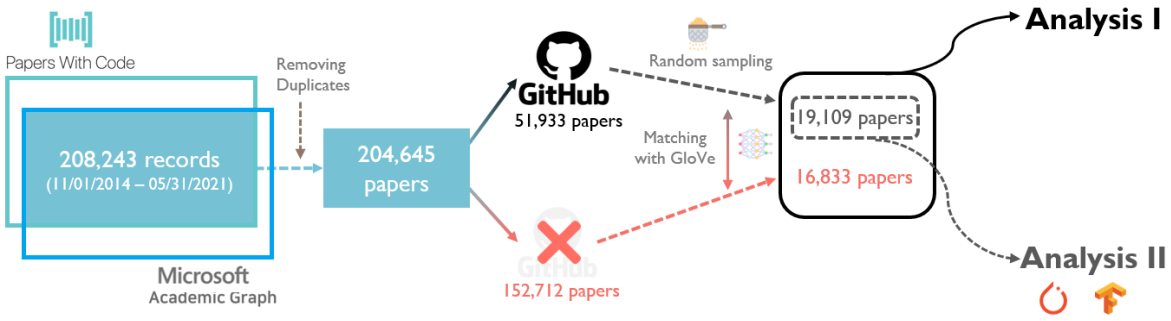
cannot be easily validated or extended because no or less popular framework is used in the implementation.

In sum, we posit that the popularity of research tools—in this context, ML frameworks—may induce an indirect network effect. We specifically hypothesize that the popularity of various ML frameworks could generate a network effect on the relevance and perceived value of individual ML research articles. This, in turn, leads to a positive impact of the popularity of an ML framework deployed in code implementation (and shared via GitHub) on the citation rate of research articles (Hypothesis 2). We test this by linking monthly shares of different ML frameworks collected by Papers with Code to the subsequent citation rates of ML research papers. Details are discussed in the later sections.

Data

The data for our analysis joins three different data sources, (1) Microsoft Academic Graph, (2) Papers with Code (PwC), and (3) repository-level metadata from the GitHub API. Figure 3.1 summarizes the data processing and the analyses we will present in the following sections.

Figure 3.1: Data Processing and Analysis Steps



Microsoft Academic Graph

We utilized the Microsoft Academic Graph (MAG) (Wang et al. 2019; Sinha et al. 2015), a large-scale database that indexes the metadata of scholarly documents and citations. MAG has been widely employed in scientometric and information science studies, such as evaluating the transdisciplinary impact of research (Y. Huang et al. 2022), characterizing career patterns of researchers and scientists (Zhao, Bu, and Li 2021), and modeling the topic selection behaviors of scientists (S. Huang et al. 2022).

This study used complete snapshot files from the December 2021 version of MAG (released on December 4th, 2021). This final official release, downloaded and housed within the data storage of the Knowledge Lab at the University of Chicago allocated by the Research Computing Center, includes approximately 270.7 million documents and nearly 19.5 billion citations. We queried MAG with arXiv identifiers and titles of ML papers indexed by *Papers with Code*. For the first analysis, we use metadata of research articles identified by MAG, including authors' information and field of study (FoS) tags assigned to each paper to match papers *with* code to those *without* code. We also leverage a feature from MAG called 'paper families,' which groups different versions of nearly identical articles appearing across various academic venues, such as preprint repositories like arXiv and conferences. This feature allowed us to trace the citation counts of a given ML research article from its very first public debut in any form.

Papers with Code

Our work used the Papers with Code (PwC) database. Maintained by the Meta AI research team and also in official partnership with arXiv since October 2020, PwC is the largest

platform linking ML research articles with their corresponding repositories (Martínez-Plumed et al. 2021). The PwC team checks code availability in GitHub[1] for every open-access ML paper and further invites community contributions. They also integrate other data sources such as NLP Progress (http://nlpprogress.com/), EFF AI metrics (https://www.eff.org/es/ai/metrics), SQuAD (https://rajpurkar.github.io/SQuAD-explorer/). The PwC dataset has been used to examine the research activities of the ML community, including benchmark practices (Martínez-Plumed et al. 2021) and dataset life cycles (Koch et al. 2021). We downloaded two snapshot files from PwC (licensed under CC BY-SA 4.0) on March 31st, 2022: 1) metadata records of 285,964 papers in the machine learning domains and 2) 90,084 papers-to-repository linkage. We additionally used the monthly framework-level popularity data identified by PwC for our analysis as detailed in our second analysis.

Our work concentrates on PwC records indexed between November 1st, 2014, and May 31st, 2021. The start date corresponds with when the PwC website provided the proportion of papers with codes. We chose May 31st, 2021, as the endpoint to ensure at least six months of paper citation windows for papers, given that the last MAG we used was released in Dec 2021. After applying this timeframe, we retained 208,243 records from PwC. Then, as detailed in Appendix, we connected these PwC records with corresponding MAG document instances using the papers' arXiv IDs and titles. We successfully matched 99.04% of 208,243 records from PwC

[1] Despite the existence of alternative platforms such as GitLab and Bitbucket, our work focuses on GitHub, the dominant platform for hosting repositories for ML papers. Accordingly, we found that most of these repositories (~99.1%) were hosted on GitHub within the PwC dataset obtained on March 31, 2022, which we use for this work. We also found that 98.7% of repositories' URLs referred to GitHub (GitLab and Bitbucket were ~0.72% and ~0.38%, respectively) in the PwC snapshots in June 2023. Nonetheless, we acknowledge the limitation of our approach and the dataset mainly focusing on GitHub, which suggests exciting directions for future research to explore the dynamics of ML research sharing on other platforms.

with MAG. This further allowed us to identify and remove duplicated records from the PwC dataset (i.e., instances where multiple PwC records were linked to the same MAG document). We were left with 204,645 unique papers recognized by PwC between November 1st, 2014, and May 31st, 2021, with 51,933 of these ML papers linked to at least one GitHub repository.

GitHub Repository Metadata

GitHub is the leading platform for hosting and distributing open-source software (Cosentino, Cánovas Izquierdo, and Cabot 2017). GitHub's popularity extended into the ML research community, with its capacity for data sharing and accessibility of model implementations based on ML papers; thus, the metadata of GitHub repositories and the activities of its users have provided an invaluable resource to empirically investigate the processes and practices of code development, sharing, and usage within the ML community (Gonzalez, Zimmermann, and Nagappan 2020; Bhattarai, Ghassemi, and Alhanai 2022; Färber 2020).

Sampling of Papers with Code

We extracted a random sample of 20,000 papers from the PwC database, all indexed between November 1st, 2014, and May 31st, 2022, and with at least one code implementation recognized by PwC. Applying the GitHub API[2] to repository URLs from PwC, we crawled the metadata of corresponding GitHub repositories, including the creation dates and repository node IDs. We note that our following analysis pivots on 19,109 papers with GitHub repositories. The reasons for exclusion were papers' actual debut dates identified by MAG were outside

[2] https://docs.github.com/en/rest

November 1st, 2014, and May 31st, 2022 range; GitHub repositories were inactive or private and thus inaccessible through API. We collected the metadata of 39,782 unique GitHub repositories linked to 19,109 papers, which we eventually used to determine the first repository for each paper by comparing the creation dates of repositories.

Analysis I: ***Impact of First Repository on Paper Citation Rate***

The primary aim of *Analysis I* is to assess the change of rate in the monthly citation after the first GitHub repositories accompanied with papers become available. Based on 19,109 papers with codes sampled as outlined above, we constructed a control group of papers *without code* by employing a nearest-neighbor matching, which we will detail in the following subsection. Subsequently, we apply the conditional fixed-effects Poisson model (Hausman, Hall, and Griliches 1984; Azoulay, Furman, and Murray 2015; Azoulay, Fons-Rosen, and Zivin 2019) to the matched samples of papers *with code* and papers *without code*, allowing us to estimate the impact of the first GitHub repositories on monthly citation rates of the papers.

Nearest-Neighbor Matching from Papers with Code to without Code

Diverse inferential methodologies have been adopted in the field of information science and management recently, such as mediation analysis applied to observational data (Díaz-Rodríguez et al. 2023; Jiang, Zhang, and Pian 2022), decomposition of time-variant effects (Xie et al. 2023) and spatial factors (Choe, Baek, and Kim 2023).

In this work, we employ a matching method, a widely used technique for constructing a control group that is statistically comparable to a treatment group, excluding the presence of treatments. This approach aims to mitigate bias arising from confounders that could affect the

estimation of treatment effects on outcomes (Angrist and Krueger 1999). We conducted nearest-neighbor matching to select a control group of papers (i.e., papers *without code* within the pool of papers indexed by PwC) comparable to 19,109 papers *with* code. We leveraged three types of information from papers for our matching process: 1) research topics of the papers captured by Fields of Study (FoS) tags assigned by MAG; 2) affiliation types of authors; 3) debut time of papers, as indicated by year-month pairs.

In line with recent information science literature harnessing textual information from scholarly documents (Cai et al. 2023; Chen et al. 2023; Zhang et al. 2022; Wang et al. 2022), our analysis leverages Field of Study (FoS) tags in MAG in our matching process to effectively construct a control group of papers. The FoS tags system delineates the research topics of a given publication with a neural network model trained on texts (e.g., titles and abstracts) and network topologies of authors, outlets, citations, and references from the entire publications indexed by MAG (Shen, Ma, and Wang 2018). This system has been adopted in bibliometric research, such as tracing the evolution of citation networks from AI and ML research (Frank et al. 2019) and modeling the topical search behavior of research scientists (S. Huang et al. 2022).

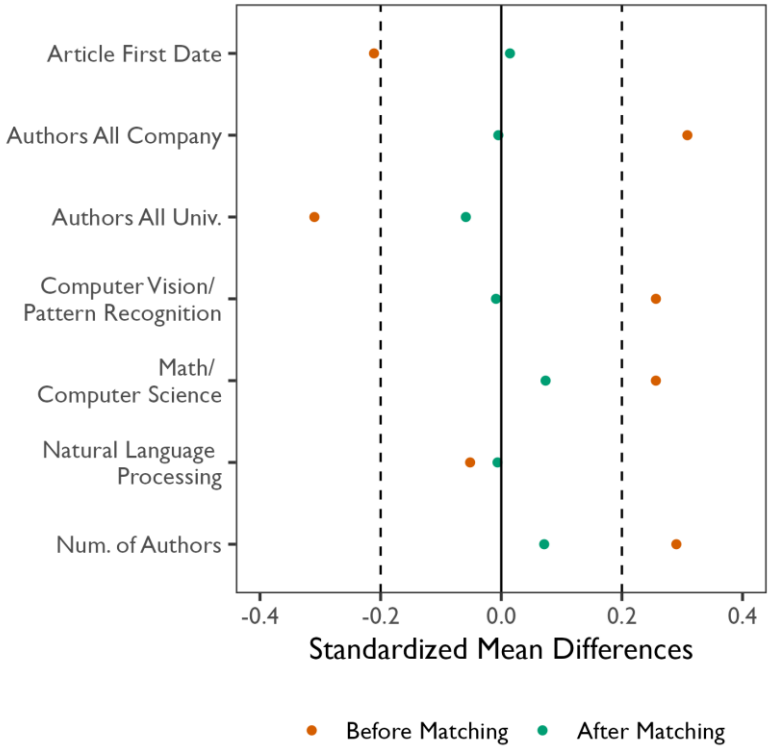
In order to characterize papers' fine-grained research topics, we first tokenized FoS tags. (the average number of FoS tags assigned to 204,645 articles is 10.38, with a standard deviation of 2.79 and a median of 10). We specifically applied the pre-trained Glove embedding model (Pennington, Socher, and Manning 2014; Young et al. 2018) to extract corresponding 50-dimensional vector representations for these tokens. Then, we computed the average values of each dimension from the token vectors, producing centroids that locate each paper's research agenda within the 50-dimensional space.

We also labeled 204,645 papers with three indicators to denote whether the list of Fos tags for a given paper includes 1) Math or Computer Science (95.2%), 2) Computer Vision/Pattern Recognition (19.6%), or 3) Natural Language Processing (12.5%). These FoS tags are directly related to these fields, as Math and Computer Science form the foundation of ML, while Computer Vision/Pattern Recognition and Natural Language Processing represent two major application areas of ML techniques. Moreover, including them makes the matching result more interpretable, as presented below.

We additionally considered the number of total authors (the average number of authors across 204,645 papers was 3.86 with a standard deviation of 2.79) and affiliation types of authors. Especially regarding the affiliation types, we labeled whether all the authors were affiliated with universities (58.4%) or all affiliated with commercial companies (3.6%).

In this way, each paper is characterized by a vector of size 56, and after standardizing 56-dimensional vector representations for papers, we selected the most similar papers without code matched to 19,109 papers with code, only when their debut dates were in the range of the same month of the same year to account for articles' birth times (Furman and Stern 2011). By doing so, we selected a control group of 16,833 papers *without code*, matched to papers *with code*. Figure 2 shows that the balance of six covariates plus the articles' first available dates from papers *with code* and *without code* improved after our matching procedure. Figure Ashows the standardized mean difference across the 50 dimensions resulting from the Glove embedding model markedly reduced after the matching.

Figure 3.2: Standard Mean Differences of Six Covariates Used in Matching Plus the First Available Dates of Research Articles



Variables and Model

Dependent Variable. We compiled our dependent variable, $MonthlyCite_{it}$ (where *i* and t represent the paper ID and the monthly age of the paper since its debut, respectively), the monthly citation counts for every 30 days from the date of the papers' debut. In cases where the Microsoft Academic Graph (MAG) contained multiple instances of the same paper (e.g., pre-prints on arXiv), we compared publication dates and traced citation trajectories from the earliest appearance of the paper in any format. Likewise, if a citing paper had multiple publication records in the MAG, we chose the earliest publication date of these records to determine the time of citation occurrence. This allowed us to comprehensively observe the monthly citation counts for each paper. Table 3.1 presents the descriptive statistics of monthly citation counts for two

groups of papers—those with an associated repository and those without— and for the combined sample.

Table 3.1: Global Distribution of Monthly Citation Count, $MonthlyCite_{it}$.

Sample Category	Mean	SD	1Q	Median	3Q	Max	# Paper (i)
Papers *with* Repo	4.39	28.46	0	1	3	3344*	19,109
Papers *without* Repo	0.54	1.84	0	0	0	162	16,833
All	2.60	20.91	0	0	1	3344	35,942

*The paper titled "Deep residual learning for image recognition" from He et al. (2015) garnered 3,344 citations between 2020-12-13 and 2021-1-12 within MAG.

For our analyses, we exclude papers that received outsized academic attention; these outliers may represent papers that introduced groundbreaking techniques or exceptionally influential algorithms in the field of ML. Including them could overshadow the underlying trends and relationships between the usability of the GitHub repository and forward citations. Thus, we attempt to minimize the impact of these outliers that potentially skew the estimation, thereby enhancing the robustness and reliability of the following analysis.

Instead of applying a single threshold to define outliers within our dataset, we employ multiple values. To do so, we first assigned cumulative citation rank percentiles measured at the end of November 2021 among the 204,645 papers linked from PwC to MAG. Then, we define papers falling within the top 0.1%, 0.5%, or 1% percentiles in the cumulative citation distribution as outliers (corresponding to cumulative citations exceeding 1,879, 604, and 370 by the end of November 2021, respectively). Tables A3.1–A3.3 in Appendix present the distribution of $MonthlyCite_{it}$ across samples without outliers of the top 0.1%, 0.5%, and 1% cited papers, respectively. With these multiple thresholds for outlier exclusion, we aim to demonstrate the robustness and consistency of our analyses.

Treatment. To estimate the change of rate in the monthly citation after papers' first GitHub repositories become publicly available, we constructed an indicator variable, $AfterRepo_{it}$. This variable switches from 0 to 1 a month after the creation of a GitHub repository linked to an article, capturing the main effect of our interest. For instance, if a repository was created four months after a paper's debut, *AfterRepo* at t = 0, 1, 2, 3 is denoted as 0, switching at t = 4 to 1 and for the subsequent periods.

To dynamically illustrate the effect of GitHub repositories, we also created $MonthDiffRepo_{it}$. It represents the differences (in months) relative to the time of the creation of the first GitHub repositories. This means that for a paper whose first repository was made available four months after the paper's debut, $MonthDiffRepo_{it}$ at t = 0, 1, 2, 3 are assigned as -3, -2, -1, 0, respectively, and at t = 4 changes to 1, then increases to 2 at t = 5, and to 3 at t = 6 ⋯ for the following periods.

Further Controls. As we trace the citation history of articles from their initial public appearance in any form, we also seek to control the effects of conference and journal publications on citation rates with two additional terms: *AfterConfPub* and *AfterJourPub*. Analogous to *AfterRepo*, these terms switch from 0 to 1 a month after publication in their respective formats.

For our estimation, we additionally account for two fixed effects: γ and $\Phi(MonthAfterDebut)$. The former, γ, denotes paper-level fixed effects, controlling for unobserved individual paper-level characteristics that may affect citation rates. The latter, $\Phi(MonthAfterDebut)$, refers to a function designed to model the life-cycle effects of articles on citation rates. To account for this effect of the ages of a given paper in a flexible manner, we use

86 indicator variables that represent the monthly ages of papers after the debut date of each article.

Model. We employed the fixed-effects Poisson model (Hausman, Hall, and Griliches 1984; Azoulay, Furman, and Murray 2015; Azoulay, Fons-Rosen, and Zivin 2019) to estimate the impact of the availability of GitHub repositories on the citation rate, as written below in Eq. (3.1). Formerly, the model can be expressed as:

$$E[MonthlyCite_{it}|X_{it}] = exp[\beta_0 + \beta_1 AfterRepo_{it} + \beta_2 AfterConfPub_{it} + \beta_3 AfterJourPub_{it} + \gamma_i + \Phi(MonthAfterDebut_{it})] \quad \cdots \text{Eq. (3.1).}$$

The coefficient for *AfterRepo*$_{it}$ (β1) holds particular interest. As introduced before, we define *AfterRepo*$_{it}$ to switch from 0 to 1, one month after the creation of the first GitHub repositories linked to papers. The coefficient, β1, captures the average effect of having a GitHub repository on the citations garnered by the corresponding paper across monthly ages of the given paper since its debut. The controls, *AfterConfPub*$_{it}$,, *AfterJourPub*$_{it}$ Φ(*MonthAfterDebut*$_{it}$) are as discussed above.

We further explore the effects of the first GitHub repositories dynamically by incorporating *MonthDiffRepo* in the estimation. In Eq. (3.2), we denote this effect as δ(*MonthDiffRepo*$_{it}$) and model it flexibly as indicator variables. And it is worth noting that while we include *AfterRepo*$_{it}$ in Eq. (3.2), it is technically omitted in estimation due to perfect multicollinearity.

$$E[MonthlyCite_{it}|X_{it}] = exp[\beta_0 + \beta_1 AfterRepo_{it} + \beta_2 AfterConfPub_{it} + \beta_3 AfterJourPub_{it} + \delta(MonthDiffRepo_{it}) + \gamma_i + \Phi(MonthAfterDebut_{it})] \quad \cdots \text{Eq. (3.2).}$$

Result

Table 3.2 presents the estimated coefficients of *AfterRepo* ($\beta 1$) across the three samples (excluding papers that received an outsized number of citations). Overall, the results suggest a positive effect of having a GitHub repository on the subsequent citation trajectories. The estimated coefficients are consistent across the three samples, each with a different threshold for removing outliers, from .211 in the first column to .198 in the third column, with statistical significance ($p < .001$). The magnitude of $\beta 1$ indicates that following the creation of their GitHub repositories, papers *with codes* had a substantial advantage in citations, from 21.9% (=exp[.198] - 1; Table 2, column 3) to 23.5% (=exp[.211] - 1; Table 2, column 1) compared to the monthly citations accrued by papers *without code*.

Table 3.2: Estimates from the Conditional Fixed-Effects Poisson Model in Eq. (3.1)

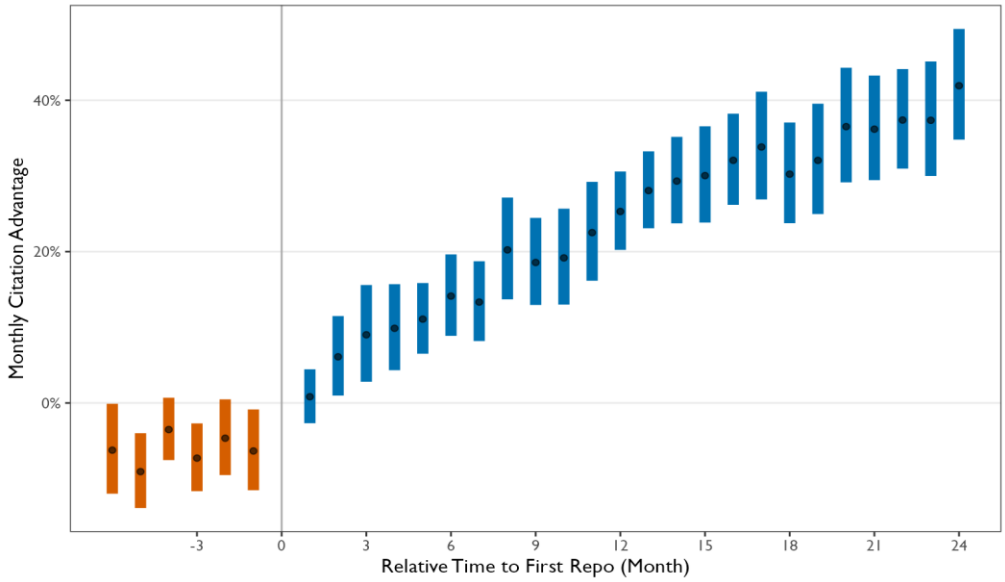
	(1) Without 0.1%	**(2) Without 0.5%**	**(3) Without 1.0%**
Dependent Variable	**Monthly Citation Counts**		
AfterRepo (β_1)	.211***	.200***	.198***
	(.018)	(.016)	(.017)
AfterConfPub (β_2)	.319***	.288***	.275***
	(.028)	(.026)	(.026)
AfterJourPub (β_3)	.314***	.307***	.325***
	(.033)	(.027)	(.027)
Num. Obs.	1,085,047	1,052,127	1,023,087
Num. Papers	35,760	35,187	34,637
Num. Month-Time	86	86	86

Note: Standard errors are clustered at individual papers and month-times. Without 0.1%, 0.5%, and 1.0% in columns mean that we excluded articles whose cumulative citations by the end of November 2021 fall under 0.1%, 0.5%, and 1% of the citation rank percentiles from all the papers cataloged in PwC and debuted between November 1st, 2014, to May 31st, 2021, for estimation.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

Figure 3.3 visualizes this effect dynamically using the samples that yielded estimates from column 2 in Table 3.2. The upper and lower ends of the bars represent the 95% confidence interval (with standard errors clustered at papers and month-time) surrounding the estimates from 6 months before and 24 months after the creation of the initial GitHub repositories associated

with the papers. (Figures A3.4 and A3.5 in the Appendix show the dynamics estimated from the samples corresponding to columns 1 and 3 in Table 3.2.) Figure 3.3 reveals no clear pre-trend in citation rates from 6 to 1 month before the earliest GitHub repositories associated with the articles become available, suggesting the balance between the treatment and the control groups. The graph also indicates that the benefit in monthly citation rates of post-first repositories escalate over time.

Figure 3.3: Dynamics Effects of Pre-and-Post First GitHub Repository on Citations



Note: The dots in the figure represent coefficients for $\delta(MonthDiffRepo)$, where δ is a function that maps the number of months after the first GitHub repositories became available into indicator variables. The interaction terms are included on top of other fixed effects and two covariates described in Eq. (3.1) using articles whose rank percentile for citation counts under 99.5% (raw cumulative citation counts by the end of November 2021 within MAG fewer than 604) corresponding to the data used for the second column in Table 2. The upper and lower ends of the bars show the 95% confidence interval (using robust standard errors clustered at papers and month-time) around the estimates.

Robustness Check

We provide estimation results applying the least square model, including the same controls and fixed effects as Eq. (3.2), to logged monthly citation counts, $ln(MonthlyCite_{it} + 1)$. The results shown in Table A3.4 in the Appendix yield similar estimations.

Analysis II - Effect of Framework Popularity on Monthly Paper Citation

In our second analysis, we examine the extent to which the popularity of ML frameworks affects the monthly citation rates of ML papers (H2). This necessitates us to focus on the citation trajectories of papers with GitHub repositories, as it is inherently impossible to estimate the impact of a framework's popularity on the citation stream of articles without repositories. In other words, this analysis investigates whether the temporal shifts in the popularity of various frameworks, such as PyTorch and TensorFlow, employed in the first repository, affect the monthly citation rates within the pool of ML papers linked to GitHub repositories.

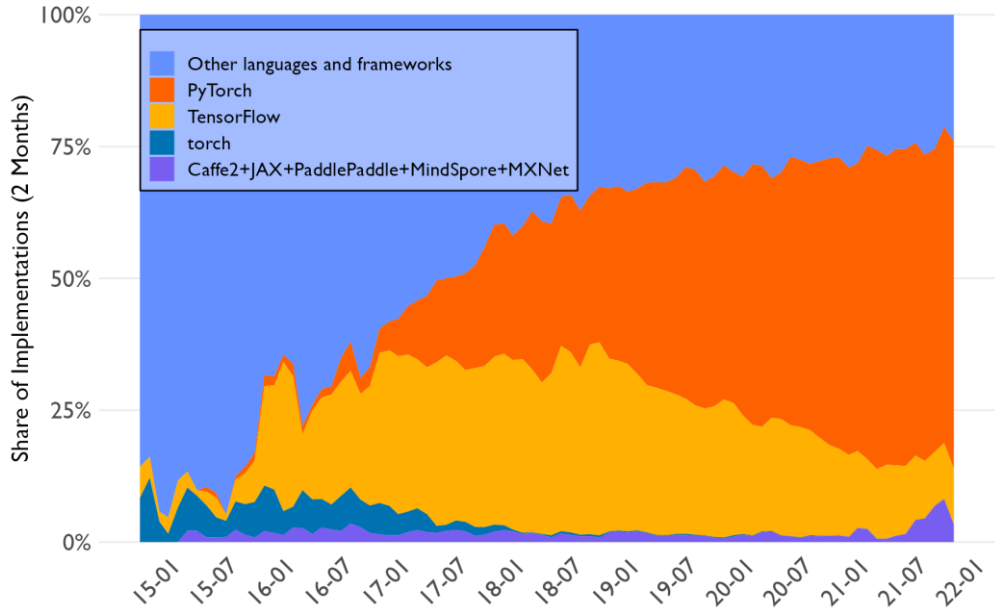
Variables and Model

Dependent variable. The definition of the dependent variable in the second analysis, $MonthlyCite_{it,}$ *remains* the same as the first analysis. However, as mentioned previously, we here concentrate on papers *with code*, the samples corresponding to the first row of Table 3.1.

Treatment. The second analysis aims to evaluate the impact of the popularity of ML frameworks employed in GitHub repositories on citation rates. To accomplish this, we retrieved the monthly shares of ML frameworks used in GitHub repositories that PwC compiles. Figure 3.4 illustrates the temporal trends of the popularity of selected ML frameworks from November 2014 to December 2021 based on the two-month rolling averages of the shares of frameworks used in GitHub repositories indexed by PwC. We calculated $FramePopularity_{it}$ after the creation of the first repositories per paper, corresponding to the type of framework used in those repositories and calendar time (i.e., the year-month pair). This allowed us to join the time-varying popularities of frameworks to our observation. For instance, $FramePopularity_{it}$ in April

2021 for a paper with a GitHub repository using PyTorch is set to be 60.54%, a two-month rolling average from the end of March (63.17%) and April (57.91%). Eventually, we interact this, $FramePopularity_{it}$, and $AfterRepo_{it}$ to investigate the impact of framework popularity on citation rates.

Figure 3.4: Two-Month Rolling Averages of Framework Shares in Code Implementation from Papers with Code



Note: The records for minor frameworks such as Caffe2, JAX are combined for visualization. Other languages and frameworks include cases in which Python was used with packages such as numpy and scikit-learn.

Note that we do not consider commonly used Python libraries, such as Numpy, Pandas, SciPy, or Scikit-learn, as ML frameworks. Although these libraries support and facilitate scientific computing, numerical operations, and even some rudimentary machine learning tasks, specialized frameworks like PyTorch and TensorFlow are designed to provide extensive support for model development, training, and deployment. These cater to the advanced requirements and needs of researchers and practitioners working on more cutting-edge areas and domains. An inherent drawback of concentrating on ML-specialized frameworks is that it may introduce a

downward bias, as the interaction term will remain zero when a specialized framework is not used in a repository even after it becomes available. Nevertheless, we posit that by adopting this conservative approach, our analysis is better positioned to assess the influence of ML framework popularity on the paper-level citation rates within the context of contemporary ML research.

Further Controls. The same set of control variables used in the first analysis is incorporated. We employ *AfterConfPub* and *AfterJourPub*, which switch from 0 to 1 one month following publication in a conference or a journal. We also use two fixed effects, γ and Φ(*MonthAfterDebut*), which represent paper-level fixed effects and month-age effects modeled with indicator variables, respectively.

Model. We used the fixed-effects Poisson regression model again to estimate the impact of the popularity of the ML framework deployed in the first GitHub repositories linked to papers on citation rates, as shown in Eq. (3.3):

$$E[MonthlyCite_{it}|X_{it}] = exp[\beta_0 + \beta_1 AfterRepo_{it} + \beta_2 AfterRepo_{it} * FramePopularity_{it} + \beta_3 AfterConfPub_{it} + \beta_4 AfterJourPub_{it} + \gamma_i + \Phi(MonthAfterDebut_{it})] \quad \cdots \text{Eq. (3.3).}$$

Result

Table 3.3 lists the estimation results for our model described in Eq. (3.3). The coefficients for *AfterRepo*FramePopularity* (β2) in Table 3.3 indicate a statistically significant advantage of approximately 0.143% to 0.157% in monthly citation rates per 1% increase in shares of a given framework. These results are consistent across the three samples, each with a different threshold for excluding outliers (equivalent to samples of the papers *with code* used in the first analysis).

Table 3.3: Estimates from the Conditional Fixed-Effects Poisson Model in Eq. (3.3)

	(1) Without 0.1%	(2) Without 0.5%	(3) Without 1.0%
Dependent Variable	**Monthly Citation Counts**		
AfterRepo (β_1)	.111***	.097***	.098***
	(.022)	(.020)	(.020)
*AfterRepo**	.144*	.156***	.153***
FramePopularity (β_2)	(.057)	(.048)	(.047)
AfterConfPub (β_3)	.295***	.263***	.249***
	(.029)	(.027)	(.028)
Num. Obs.	573,289	541,697	514,910
Num. Papers	18,929	18,375	17,861
Num. Month-Time	86	86	86

Note: Standard errors in parentheses are clustered at individual papers and month-times.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

Although the unit magnitude of the effect size may appear modest at best or even trifling at worst, considering the recent gap between the two frameworks, PyTorch and TensorFlow (as depicted in Figure 3.4), provides a more tangible sense of the effect. For example, in April 2021, the two-month rolling averages for the shares of code implementations in PyTorch and TensorFlow were 60.54% and 13.19%, resulting in a difference of 47.35%, according to PwC data. When this gap is applied to the estimated model based on the sample excluding the top 0.5% cited papers (column 2 in Table 3), it translates to a 7.67% (= exp[.156*.4735] - 1 ≈ 0.0767) (dis) advantages in the monthly citation rate associated with the type of framework used in the first code implementation. We believe that such an effect deserves attention. And we note again that our estimation may be conservative because of the focus on specialized ML frameworks.

Robustness Check

Analogous to *Analysis I*, we provide the estimation results using the least square model, incorporating the same controls and fixed effects for the logged monthly citation counts. The

resulting estimates for *AfterRepo* FramePopularity (β*2), as shown in Table A3.5 in Appendix are positive once again.

Discussion and Conclusion

The growing emphasis on research transparency has led ML researchers to share their model implementations via public repositories. In this study, we aim to assess the extent to which ML research articles gain advantages in citation rates after their GitHub repositories become available. Additionally, we investigate the relationship between the popularity of ML frameworks used in code implementation and the citation rates of ML research papers, drawing upon the theory of network effects and cognitive shortcuts. We conducted a linkage across Papers with Code (PwC), Microsoft Academic Graph (MAG), and repository-level metadata collected through GitHub API. Our findings reveal that ML papers experience a significant increase of approximately 20% in monthly citation rates following the creation of their first accompanying GitHub repositories, compared to papers without such repositories. Furthermore, our analysis demonstrates that the popularity trends over time for different ML frameworks employed in these initial repositories exhibit a second-order network effect, which influences the monthly citation rate of the associated research articles.

Our study provides several implications. First, aligned with the call from the *Open Science* movement for research transparency, our results highlight that the extra efforts researchers put into preparing replication materials, whether required by journals and conferences or not, is not in vain. These endeavors can potentially increase academic impact when researchers make such scientific artifacts publicly accessible. However, our result should not be interpreted as encouraging researchers to devote intense amounts of time and effort to

establish and sustain a repository purely for the sake of visibility. Instead, we suggest maintaining clear documentation and shareable implementation throughout research projects, so researchers save time and cognitive resources. Our result implies that this not only helps them to respond to the institutional desideratum but also to benefit themselves. Relatedly, it is also worth noting that while common for corporations to showcase their technological prowess in academic venues, decisions at the firm-level regarding the extent to which they follow the principle of open-source software (and science) may require striking the delicate balance between the protection of core technology and the pursuit of increased credibility, impact, and reputation—all of which should be in line with strategic objectives (McIntyre and Srinivasan 2017).

Second, our analysis provides broader implications. Take, for example, the recent proliferation of Large Language Models (LLMs) following the release of the ChatGPT interface in late November 2022. Within less than a year, a wave of functionally similar and comparable models and services, such as Llama, Alpaca, Langchain, Pinecone, have emerged, each vying for a stake in the domain. Our result provides insights into how this LLM landscape might evolve; researchers, practitioners, and organizations may stand to enhance their visibility and impact by aligning their work with more popular and widely supported and endorsed models However, as the competition intensifies, researchers and developers may find that it is increasingly crucial to keep pace with the most popular and widely adopted tools to ensure their work remains relevant and accessible.

Third, while our analysis underscores the advantage of utilizing widespread technological infrastructure in disseminating research outputs, the contrasting disadvantages stemming from deploying less popular technological infrastructure deserve attention. Historical precedents, such as the widespread acceptance of the QWERTY keyboard layout, remind us that technical

consensus does not necessarily equate to progress or optimality but can be viewed as an outcome of path dependency. This perspective suggests that innovative research and ideas may face challenges if built on less prevalent technological platforms, despite their advanced technical capability and value. Hence, our results also suggest caution against a technological monoculture for a research community, despite the benefits of consensus in reducing individual and collective cognitive burdens.

Our study is not without limitations. First, we acknowledge that the data used in this study is anchored to ML research papers indexed by Papers with Code (PwC). Despite its public accessibility and comprehensive coverage, prior research (Martínez-Plumed et al., 2021) noted that PwC tends to be inclusive for recent articles posted on arXiv and proceedings published in major ML/AI venues and conferences, such as the International Conference on Learning Representations (ICLR), Neural Information Processing Systems (NIPS), and the Conference on Computer Vision and Pattern Recognition (CVPR) than less prominent venues. As such, the estimated effect sizes may vary when a different data source is used. However, we maintain that this would not necessarily invalidate the overall findings of this study, considering the importance of these venues within the contemporary ML research community.

Second, the scope of the analysis is limited to GitHub repositories rather than other platforms such as BitBucket and GitLab. While the choice admittedly reflects the prevalence of GitHub within the PwC dataset, we acknowledge the potential limitations introduced due to this focus. Distinct features and interfaces other platforms provide could influence how research is disseminated. While our analyses provide meaningful insights into the role of technological infrastructure on research diffusion, we believe that future studies can explore the heterogeneous effects of different hosting services on research diffusion. For instance, we could enrich the

understanding of the multi-layered network effects created through the interactions between the types of repository hosting services and ML frameworks.

Third, it is crucial to note that the changes in the popularity of frameworks primarily manifest through the shifts in code implementation shares between TensorFlow and PyTorch. The record from PwC indicates a smooth handover from TensorFlow to PyTorch in the ML research community without abrupt transitions (at the time of this writing, TensorFlow still maintains a substantial user base). As our observation may primarily reflect the bilateral dynamic between the two major frameworks, we admit that competitions involving more than a dyad may yield different outcomes.

Lastly, our analysis does not explicitly distinguish between long-term and short-term effects, nor does it consider the variations in the prestige of different academic venues. For instance, Kwon & Motohashi (2021) showed that papers disclosing data attract more citations in the short term, but the advantages often wane over time; and this tendency is particularly pronounced for papers published in low-reputation journals. Our study design that hinges on paper-level fixed effects admittedly makes it challenging to explore this dynamic thoroughly. However, we note that these represent exciting directions for future research to paint a more nuanced picture of scholarly communication.

Scientific artifacts beyond research manuscripts—particularly code repositories—have gained increasing importance. By combining metadata of ML research articles and code repositories, this study addressed the enduring issue that academic impact could be shaped not only by the merit of research but also by various factors. While we do not conclusively adjudicate why researchers may choose (or not) to post their code implementation and use specific ML frameworks, our analysis provides insights into the unintended consequences amid

intense competition. As the research community continues to embrace the principle of *Open Science*, it is critical to understand the complexities and challenges posed by these dynamics to more effectively support and balance different objectives of research transparency, innovation, and scientific rigor.

Appendix for Chapter 3

Data Linkage between PwC and MAG

Matching PwC records with arXiv ID. Out of 208,243 PwC records, 174,265 had arXiv IDs. We used this information to query the PaperURL table in MAG with two conditions: a URL containing 1) "arxiv.org" and 2) the queried arXiv ID. To avoid false matches like '1606.0365' being matched to '1606.03657', we double-checked if returned URLs contained the exact same queried arXiv ID. This step matched 99.94% of records with arXiv IDs to at least one MAG paper ID, leaving 108 unmatched. For the remaining 107 records, we queried MAG papers' raw titles using PwC-indexed raw titles, matching 62 records and leaving 45 unmatched. For these 45 records, we normalized PwC paper titles, removed whitespaces and non-alphanumeric characters, and finally converted texts to lowercase. Then, we queried the MAG Papers table using MAG-provided normalized titles, capturing MAG paper IDs for 9 out of 45 records.

Matching 33,978 PwC records without arXiv ID. For 33,978 PwC records without arXiv ID, we queried the MAG PaperURL using two URL types from PwC records: one for abstracts and another for PDF files. This step linked 5,420 records to at least one MAG paper ID, leaving 28,558 unmatched. For these 28,558 records, we conducted a raw title search similar to A1, linking 22,670 records to MAG paper IDs and leaving 5,888 unmatched. Using title normalization and querying for the remaining 5,888 records, we connected 3,553 to at least one MAG paper ID, leaving 2,335 without linkage.

***Fuzzy title matching for unmatched records*.** Steps in A1 and A2 left 2,371 PwC records without MAG paper ID linkage (36 after A1 plus 2,335 unmatched records after A2). For these

records, we applied fuzzy title matching. We first restricted the search space using MAG's field of study tags: 'Artificial intelligence,' 'Computer vision,' 'Pattern recognition,' 'Machine learning,' 'Natural language processing,' and 'Artificial neural network.' We compared the normalized MAG publication titles with the PwC-identified publication within the same year and calculated the Levenshtein distance similarity score. With a threshold of 90, we flagged matching candidates, capturing 373 records' MAG titles. This method identified cases where minor typos prevented exact raw or normalized title matching, such as "weakly supervised deep functional maps for shape matching" in PwC to "weakly supervised deep functional map for shape matching" in MAG. By performing the multiple steps of record linkage from A1 and A3, we link the 206,245 (99.04% of 208,243 PwC records between Nov 1st, 2014, and May 31st, 2021) to at least one MAG paper ID.

Inspecting duplicates. After the initial linkage between the PwC records to MAG paper IDs, we examine record duplication within PwC records matched to MAG paper IDs. Among 206,245 PwC records, we found 2,789 had overlapping MAG paper IDs with others. We iterated through the 1,569 MAG paper IDs and preserved one of multiple PwC records sharing MAG paper IDs by prioritizing records matched through arXiv ID (A1) and selecting the earliest records in the PwC database when arXiv ID was not available. This process removed 1,398 out of 206,245 PwC records, keeping 204,847. At this stage, the number of unique MAG paper IDs linked to 204,847 PwC records was 272,037.

Update MAG paper IDs by MAG paper family. MAG defines a “paper family” when the same paper appears in multiple venues such as a repository (e.g., arXiv), conference proceedings,

and journals. In this case, MAG assigns unique paper IDs for each publication but designates one of the papers as a *primary publication* (called "family ID"). We leveraged this feature to update the MAG paper IDs linked to PwC records to include all relevant IDs. We queried the primary publications' MAG paper IDs based on the papers captured so far, updating the PwC-MAG matching result. This process updated 49,152 PwC records' MAG paper ID sets, expanding the number of unique MAG paper IDs from 272,037 to 321,363. We then queried the list of all publications associated with those primary papers and updated the MAG paper IDs linked to the PwC records again. This step updated 5,233 PwC records' MAG paper ID sets, increasing the number of MAG paper IDs from 321,363 to 327,135.

Double-check PwC records sharing MAG paper IDs and remove duplicates. After updating the PwC records to MAG paper IDs linkage, we reexamined the linkage expansion for additional duplicate records, as in A4. Among 204,847 PwC records, we find 409 of them have overlapping MAG paper with other PwC records (through 419 MAG paper IDs). Typical duplicate records included cases where conference proceedings or journal article titles slightly differed from those posted in repositories. For example, "Bilinear CNNs for Fine-grained Visual Recognition" was published as "Bilinear CNN Models for Fine-Grained Visual Recognition" in ICLR 2018. (Note that those duplicate records were only identifiable after updating MAG paper ID through MAG family ID). Iterating through the 419 MAG paper IDs causing duplicates, we selected one of multiple PwC records sharing MAG paper IDs, prioritizing records matched through arXiv ID. This resolved 409 cases with overlapping MAG paper IDs. For the remaining 10 cases, we manually checked duplicates. In the end, we removed 202 PwC records. Finally, we

confirmed that none of the remaining 204,645 PwC records shared MAG paper IDs by ensuring each 327,131 MAG paper IDs appeared only once across 204,645 PwC records.

Authors' Affiliation Type Assignment. With 327,131 MAG paper IDs mapped to 204,645 PwC records, we queried MAG's Paper Author Affiliations table and extracted 1,282,249 records of MAG paper ID - author ID - affiliation ID - Affiliation Name. Initially, we used MAG's lookup table (*Affiliation.txt*) linking affiliation ID to Global Research Identifier Database (GRID) ID and their affiliation types, connecting 801,455 of 1,285,582 records to GRID affiliation types ('Archive,' 'Company,' 'Education,' 'Facility,' 'Government,' 'Healthcare,' 'Nonprofit,' 'Other'), leaving 480,794 records' affiliation types unidentified.

Among the 480,794 unidentified records, 119,457 records contained institution names, while 361,337 records had no further affiliation information directly available from MAG. To impute affiliation types for these records, we employed rule-based dictionaries to map institution names to types (Table A3.1), applied to 1) original full affiliation names and 2) normalized affiliation names from MAG, mapping 65,659 records to affiliation types.

For the remaining 361,337 records without affiliation names, we extracted publications associated with a given author ID and filtered them within two years of the focal paper's publication date (i.e., 365 days <= publication date difference <= 365 days). We then recursively extracted the paper-author-affiliation table with the collected paper IDs and imputed the given author IDs' affiliation ID and GRID affiliation type with the record with the smallest publication date difference (avg. publication date difference: 51.04 days, standard deviation: 74.53). This assigned affiliation to types for 247,926 records.

Within the two-year range, we applied the same institution names-types dictionaries for records with institution names available but without GRID affiliation types, connecting 24,454 records to affiliation types. This left affiliations for 88,957 records of paper ID - author ID (6.94% of 1,282,249) unavailable due to a lack of affiliation information in MAG or unidentifiable affiliation types based on our dictionary. However, this still suggests that we successfully assigned affiliation types to approximately 93% of 1,282,249 records.

Table A3.1: Affiliation Types for Keywords

Affiliation Types	Keywords
Nonprofit/Company	("alibaba" *or* "bell labs" *or* "bloomberg" *or* "bytedance" *or* "company" *or* "corp." *or* "corporation" *or* "deepmind" *or* "didi" *or* "disney" *or* "dji" *or* "facebook" *or* "google" *or* "ibm" *or* "inc" or "jd.com" *or* "kakao" *or* "lg" *or* "limited" *or* "linkedin" *or* "llc" *or* "ltd" *or* "lyft" *or* "mitsubishi" *or* "jp morgan" *or* "naver" *or* "netflix" *or* "nokia" *or* "openai" *or* "salesforce" *or* "samsung" *or* "sense time" or "sensetime" *or* "sony" *or* "spotify" *or* "uber" *or* "walmart" or "allen institute" *or* "foundation")
Education	("academy" *or* "college" *or* "department" *or* "dept."" *or* "polytechnique" *or* "school" *or* "u of" *or* "u. of" *or* "univ")
Government	("national" *or* "federal" *or* "nasa")
Healthcare	("hospital")

Table A3.2: Distribution of Monthly Citation Count, $MonthlyCite_{it}$, Excluding Outliers - 0.1%

Sample Category	Mean	SD	1Q	Median	3Q	Max	# Paper (i)
Papers *with* Repo	2.68	6.23	0	0	3	198	18,929
Papers *without* Repo	0.53	1.66	0	0	0	89	16,831
All	1.67	4.79	0	0	1	198	35,760

Table A3.3: Distribution of Monthly Citation Count, $MonthlyCite_{it}$, Excluding Outliers - 0.5%

Sample Category	Mean	SD	1Q	Median	3Q	Max	# Paper (i)
Papers *with* Repo	1.81	3.55	0	1	2	77	18,375
Papers *without* Repo	0.50	1.42	0	0	0	52	16,812
All	1.17	2.81	0	0	1	77	35,187

Table A3.4: Distribution of Monthly Citation Count, $MonthlyCite_{it}$, Excluding Outliers - 1.0%

Sample Category	Mean	SD	1Q	Median	3Q	Max	# Paper (i)
Papers *with* Repo	1.43	2.73	0	0	2	75	17,861
Papers *without* Repo	0.47	1.29	0	0	0	52	16,776
All	0.96	2.19	0	0	1	75	34,637

Table A3.5: Estimates from Fixed-Effects the Least Square Model for Eq. (3.1)

	(1) Without 0.1%	(2) Without 0.5%	(3) Without 1.0%
Dependent Variable	**Monthly Citation Counts**		
AfterRepo (β_1)	.263***	.218***	.187***
	(.017)	(.015)	(.013)
AfterConfPub (β_3)	.237***	.204***	.183***
	(.018)	(.015)	(.014)
AfterJourPub (β_4)	.102***	.100***	.101***
	(.012)	(.010)	(.010)
Num. Obs.	1,085,047	1,052,127	1,023,087
Num. Papers	35,760	35,187	34,637
Num. Month-Time	86	86	86

Note: Standard errors are clustered at individual papers and month-times. Without 0.1%, 0.5%, and 1.0% in columns mean that we excluded articles whose cumulative citations by the end of November 2021 fall under 0.1%, 0.5%, and 1% of the citation rank percentiles from all the papers cataloged in PwC and debuted between November 1st, 2014, to May 31st, 2021, for estimation.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

Table A3.6: Estimates from Fixed-Effects the Least Square Model for Eq. (3.3)

	(1) Without 0.1%	(2) Without 0.5%	(3) Without 1.0%
Dependent Variable	**Monthly Citation Counts**		
AfterRepo (β_1)	.008	-.002	-0.009
	(.011)	(.010)	(.010)
*AfterRepo**	.362***	.337***	.320***
FramePopularity (β_2)	(.034)	(.033)	(.033)
AfterConfPub (β_3)	.226***	.187***	.165***
	(.020)	(.016)	(.015)
Num. Obs.	573,289	541,697	514,910
Num. Papers	18,929	18,375	17,861
Num. Month-Time	86	86	86

Figure A3.1: Standard Mean Differences of 50 Embedding Features from Glove Embedding Model Used for Matching, Before and After

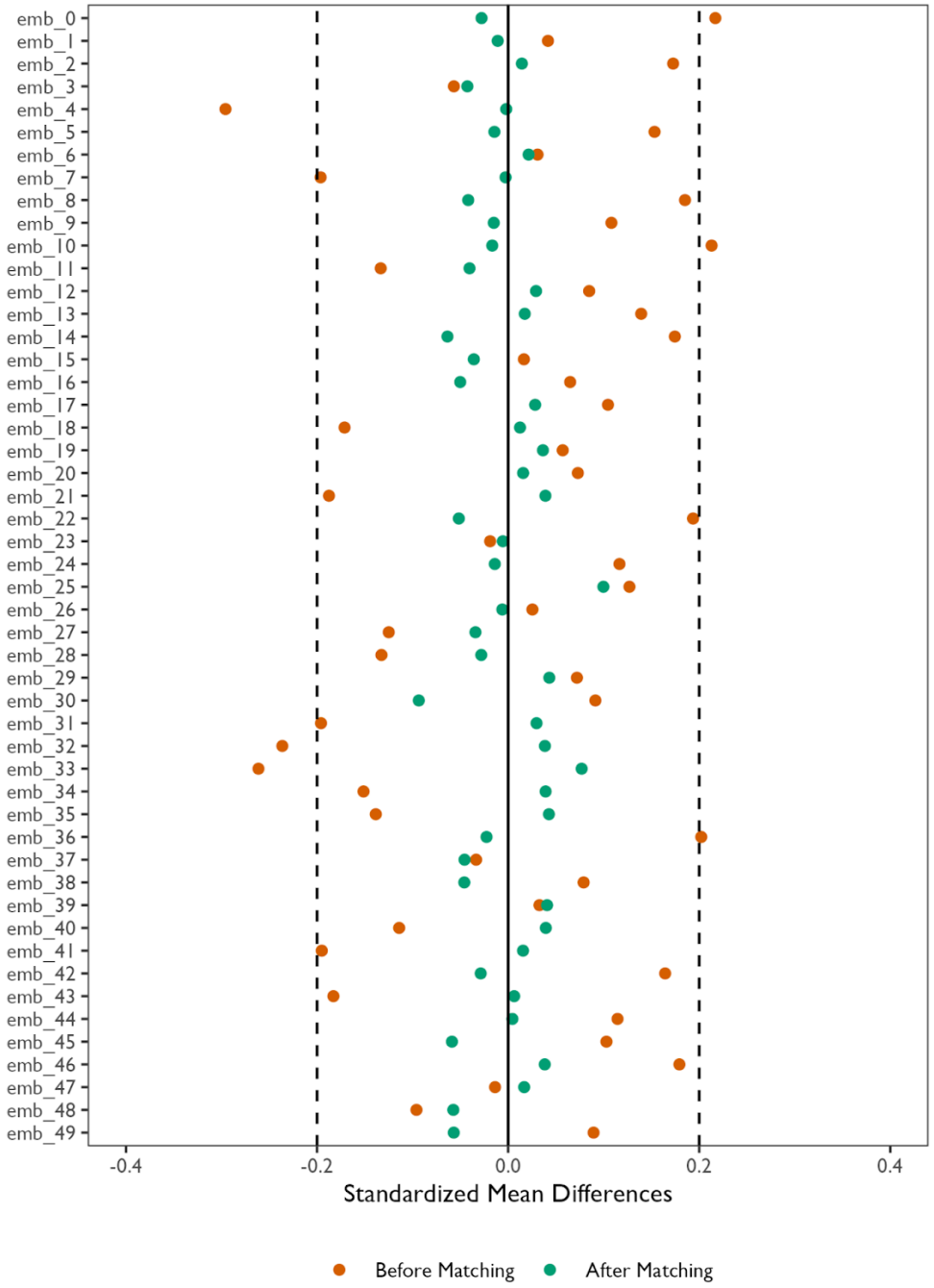
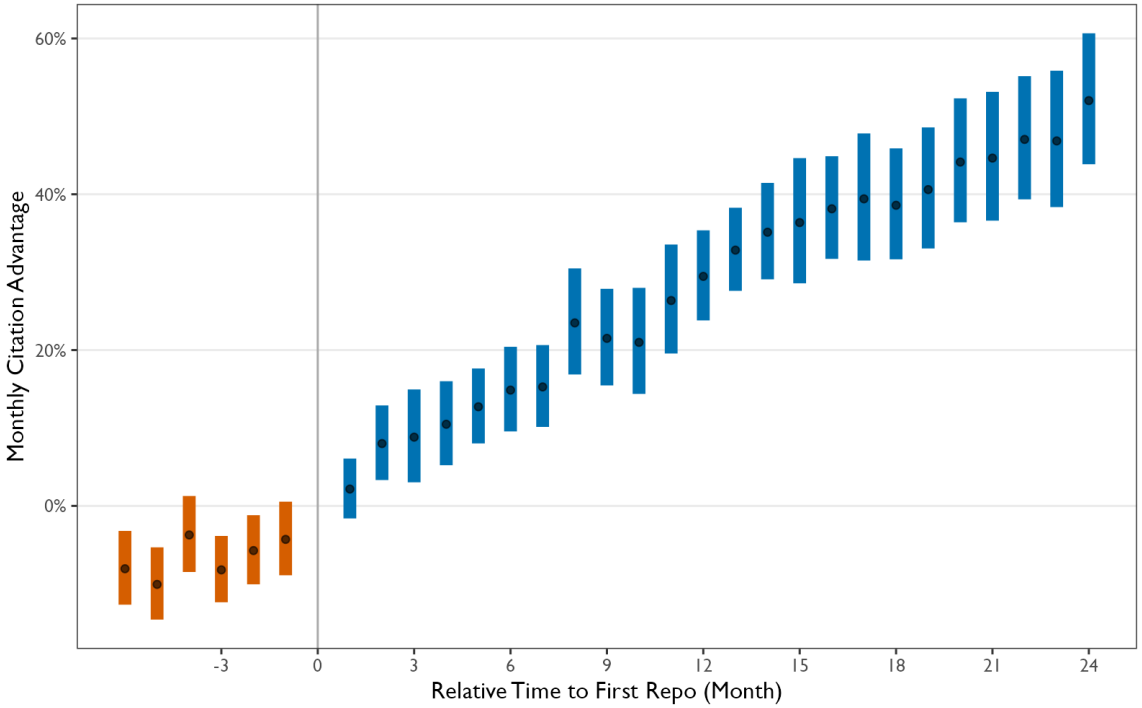
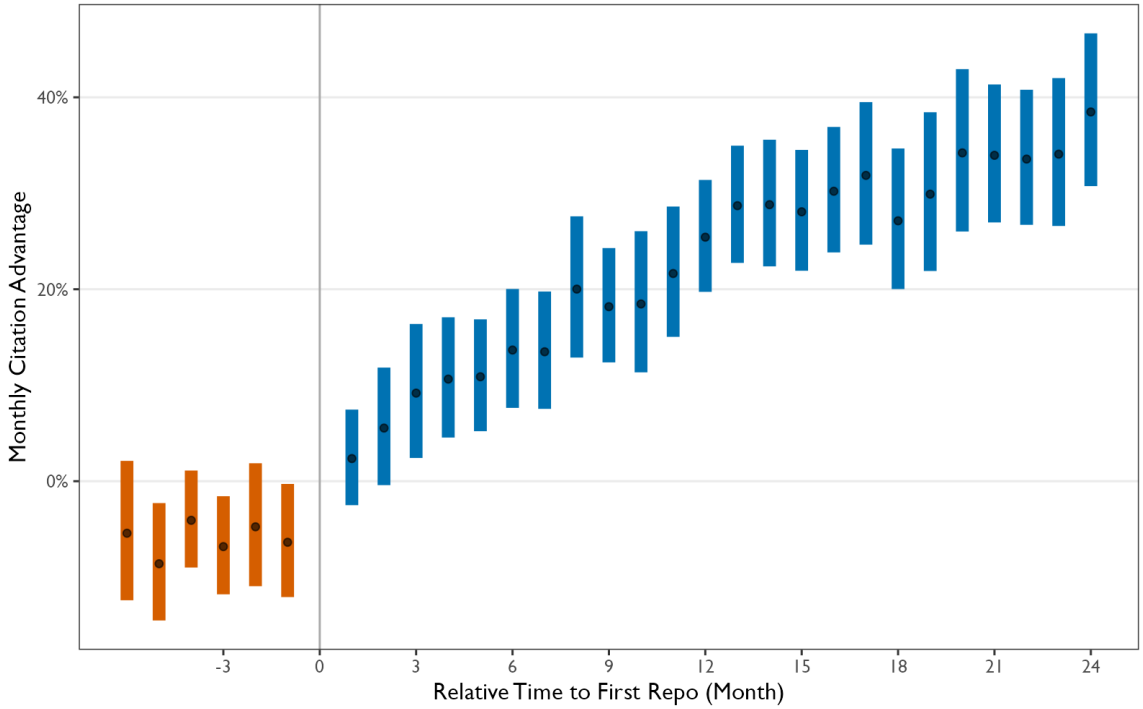


Figure A3.2: Dynamics Effects of Pre-and-Post First GitHub Repository on Citations after Excluding Articles with Rank Percentile for Citation Counts above 0.01%



Note: The dots in the figure represent coefficients for interaction terms $\delta(MonthDiffRepo)$, where δ is a function that maps the number of months after the first GitHub repositories became available into indicator variables. The interaction terms are included on top of other fixed effects and two covariates described in Eq. (3.1) using articles whose rank percentile for citation counts under 99.9% (raw cumulative citation counts the end of Nov within MAG fewer than 1,879), which corresponds to the data used for the first column in Table 3.2. The upper and lower ends of the bars show the 95% confidence interval (using robust standard errors clustered at papers and month-time) around the estimates.

Figure A3.3: Dynamics Effects of Pre-and-Post First GitHub Repository on Citations after Excluding Articles with Rank Percentile for Citation Counts above 1%



Note: The dots in the figure represent coefficients for interaction terms $\delta(MonthDiffRepo)$, where δ is a function that maps the number of months after the first GitHub repositories became available into indicator variables. The interaction terms are included on top of other fixed effects and two covariates described in Eq. (3.1) using articles whose rank percentile for citation counts under 99% (raw cumulative citation counts by the end of Nov 2021 within MAG fewer than 371), which corresponds to the data used for the third column in Table 3.2. The upper and lower ends of the bars show the 95% confidence interval (using robust standard errors clustered at papers and month-time) around the estimates.

Conclusion

Summary and Reflective Notes

Chapter 1

Chapter 1 conceptualizes the circulation of potentially latent sources of variation as "social-epistemic bubbles." This conceptualization aims not only to transcend the conventional discussion of tacit knowledge as know-how residing within individuals (Collins 2010, 2007) but, more importantly, to envision the field of science as a continuous space. Tacit knowledge in this space is neither required to be ubiquitous across the collective nor confined to localized networks of scientists, as suggested by the "Invisible College" (Crane 1972). In this vein, Chapter 1 posits that overlapping socio-epistemic bubbles can span individuals, any single community, and across communities.

The chapter maps the continuous space of socio-epistemic bubbles by applying a neural embedding model to a large-scale network data of author-publication. The resulting embedding model generates similar vector positions to authors who frequently co-author papers, as well as to articles co-authored by overlapping collaborators sharing considerable tacit knowledge. More importantly, as an autoencoder, the network embedding model locates two researchers adjacently, even if they have never collaborated directly but only share a common principal investigator or collaborate through other principal investigators. Conversely, a single large-scale study conducted by otherwise unconnected scientists would be represented in a less dense region of the space, indicating potentially low levels of tacit knowledge sharing. With this approach, I attempt to bridge network embedding models for a representational space of scientific collaborations and insights derived from the literature of social studies of sciences.

The association between social density, measured within the social embedding space, and

the outcomes reported in scientific publications is evaluated by analyzing RCTs that were collected, filtered, and clustered within Cochrane Systematic Reviews for meta-analysis, one of the most authoritative sources in evidence-based medicine (Au 2021; Salandra, Criscuolo, and Salter 2021). This empirical setting represents a conservative setting to test the proposed hypothesis — there is systematic association between the proximity of RCTs measured in the continuous social space and their estimates. The analysis supports this proposition. Nonetheless, it is important to emphasize that the intention is not to cast unsubstantiated skepticisms against RCTs or the practice of meta-analysis. As discussed in the conclusion section of Chapter 1, the findings aim to elicit conversations to explicate tacit elements, such as a more precise scope conditions under which an experiment was conducted.

This raises the question of the extent to which experimental conditions can be codified—reminiscent of the debate between Collins (1985) and Franklin (1989) on "experimenters' regress," particularly within the context of gravitational waves and physics. Their debate eventually led to a mutual acknowledgment that scientific claims are, in principle, not an infallible enterprise and that social elements substantially play a role in scientific endeavors. However, this does not necessarily have to invoke global skepticism towards scientific practice (Franklin and Collins 2016). Nonetheless, this agreement centers on validation and justification of scientific claims, without necessarily resolving many of the lingering questions on what remains tacit during the creation stage of scientific knowledge (Knorr-Cetina 1981).

This issue is particularly pronounced in experiments involving many environmental and natural variations in social, behavioral, biomedical, and health sciences. These often require experimenters to devise operationalizations that allow modification of relevant experimental parameters (Feest 2016) and also involve implicit rules that "do not contain rules for their own

application" (Collins 1985, 14). A methodological approach to address this has been randomization and, again, , mustered to synthesize estimates through meta-analysis techniques. However, as demonstrated by conflicting conclusions from meta-analyses on the association between violent games and behaviors (Hilgard, Engelhardt, and Rouder 2017; Anderson et al. 2010; Kepes, Bushman, and Anderson 2017), even verdicts from this approach do not necessarily guarantee to resolve scientific controversies, yet evidencing tacit sources of disagreements (Vrieze 2018).

Findings reported in Chapter 1 do not necessarily affirm the "strong" version of tacit knowledge, as initially proposed by Collins (1985). Instead, the implications from Chapter 1 resonate with the call for "operational analysis" that Feest (2016) advocates. This approach suggests that scientists can productively leverage uncertainty about experimental results to scrutinize and uncover the underlying tacit assumptions. The methodology introduced in Chapter 1 offers a means to identify specific loci, socio-epistemic bubbles, to prompt this type of investigation, which I hope to be fruitfully further exploited. However, this does not imply again that I recommend all such bubbles should be collapsed; this would eliminate reservoirs of diverse, tacit perspectives and conditions. As highlighted in the concluding remarks of Chapter 1, maintaining a balance is critical, as not only 'overdispersion' but 'underdispersion' or insufficient variability may present challenges to achieving robust science.

Lastly, I suspect that one would obtain comparable or more salient results if RCT-based social and policy science studies are analyzed. I posit several reasons for this likelihood. Not to mention, laboratory experiments conducted in social psychology have historically heavily depended on well-educated Western populations. More importantly, a significant portion of field experiments in social science have been conducted in villages in Sub-Saharan African countries.

Of course, these locations are chosen for various reasons, including targeting populations to improve poverty, health, and educational outcomes as part of the quest for causality under the banner of the 'randomista' movement. However, as a Nobel-laureate economist and a proponent of the randomista, Esther Duflo, once noted, "Randomized controlled trials are challenging to execute because they require rapid decision-making, and these decisions carry consequences" (Stoughton 2022) It suggests that those quick decisions made may remain tacit within socio-epistemic bubbles. Of course, the implication warrants cautious interpretation as the intention here is not to foster anti-RCT cynicism but to encourage reflection.

Chapter 2

Chapter 2 extends the concept of "bubbles" in science, likening them to bubbles pronounced in financial and asset market bubbles with inflated attention. This analogy is grounded in several key considerations. Although the chapter introduces two high-profile cases—cardiac stem cell research and cancer immunotherapy—one of the essential motivations aligns with concern about the dominant measure to determine scientific work's importance and impact (Fortunato et al. 2018; Partha and David 1994). This, in turn, undermines the relationship between short-term popularity and long-term significance. Accordingly, one critical aim is to introduce weighting methods that can better reflect the nuances in citations, which have traditionally been uniformly accounted for in bibliometric databases.

To this end, the weight/distance between two papers linked through citations was measured with two embedding spaces trained on the PubMed Knowledge Graph (Xu et al. 2020). The first is the 'social space,' the same model employed to measure the "socio-epistemic bubbles" in Chapter 1. The second space, 'scientific space,' aims to capture the direct and

indirect associations of Medical Subject Headings (MeSH) and vector positions of publications associated with MeSH combinations. While the need for distinguishing such spaces—such as the differentiation of 'cultural' and 'structural' holes (Pachucki and Breiger 2010)—has been acknowledged, direct and large-scale measurement of these differing spaces has seldom been attempted to my knowledge. But it is important to note that these spaces—those of people and ideas—are here considered distinct, yet not orthogonal; instead, they are oblique to each other.

The main analysis applies this measurement scheme to 28,504 unique biomedical subfields (Azoulay, Fons-Rosen, and Zivin 2019), relating the degree of social and scientific diffusion with a drastic decline in the given subfield's popularity. This decline is operationalized by considering the first-order difference of the standardized citation counts within a given subfield across two-year intervals, falling below extreme cutoffs. As the study title suggests, the result shows that limited diffusion anticipates collapses of biomedical subfields, highlighting the importance of tracing pathways through which scientific ideas are disseminated within and across the scientific community.

An immediate and valid criticism of this study would be that the analogy of financial bubbles might not seamlessly apply to the realm of science; analogies can be overstretched. This can particularly be the case given the premise that underlying book values support the market values of financial assets, whereas the fundamental values of scientific inquiry and outcomes are less certain. Nonetheless, resources channeled into scientific research represent social investments, yielding research outputs primarily in the form of publications. In this sense, Parallels between scientific idea markets and financial markets have been repeatedly drawn. For instance, Goldman and Shaked (1991, 31) noted: "*scientific agents act in some ways like vendors, trying to sell their findings, theories, analyses, or arguments to an audience of*

prospective buyers." Similarly, philosophers of science, Pedersen and Hendricks (2014, 507), in their article entitled "Science Bubbles" stated, *"Science, like any other investment, is prone to bubble formations and overinvestment…Overly optimistic investments in specific areas of research, methodologies, and technologies generate states comparable to the ones financial markets experience prior to crashing.*" The chapter links these propositions to the observation that the impact or return in science has been predominantly evaluated by the degree of attention they garner, typically quantified through citation counts. This reasoning guided the study to conceive that inflated attention, or bubbles, within the scientific community can be identified by examining whether certain research areas attract a disproportionate amount of attention, lacking a comparable level of genuine diffusion, and subsequently if collective attention sharply declines with corrected perception.

This criticism led to a series of post hoc analyses through which I examined further relevant patterns associated with these collapses to bolster the analogy. First, I explore the potential impact imposed by the concentration of "scientific capital" (Bourdieu 1975) by showing that the likelihood of subfield collapse is positively associated with 1) the proportion of papers authored by superstar scientists within the subfield and 2) the fraction of subfield funding accounted for by collaborators of these star scientists. Furthermore, the analyses show that collapsed subfields are collectively less clinically translatable than those that have not collapsed but garnered disproportionately more citations than expected. Additionally, authors who published their articles close to the collapse were significantly less productive 5 or 10 years after the collapse compared to those who entered the field earlier, reminiscent of investors losing money after buying assets at peak prices during bubbles.

I believe the findings may inform how the scientific community, policymakers, and funding bodies conceptualize and measure the impact and value of research. By highlighting the limitations of current bibliometric indicators and the potential for inflated attention to distort the scientific landscape, the result, on the one hand, underscores the importance of multidimensional assessment criteria. This reconceptualization could lead to a more effective allocation of research funding, encouraging investment in areas that may lead to long-term contributions to knowledge rather than short-term visibility. Moreover, by fostering a diverse and pluralistic understanding of scientific worth, policymakers and funding agencies can better support emerging fields in scientific research by discounting immediacy of attention.

Lastly, the issue of circularity may be at this point unavoidable, even with the addition of further evidence from data analysis that aligns with the initial analogy of attention bubbles. This challenge is further compounded by reliance on a newly proposed measurement scheme. Such circularity, though, is not an anomaly but a common occurrence at the nascent stages of any measurement system development. For example, as Chang (2004) persuasively illustrates, the fixed points in early thermometry—like the freezing and boiling points of water—relied on observable phenomena that were assumed to be consistent. However, the definition and measurement of these points were contingent on the very notion of temperature scientists were trying to establish.

Overcoming this challenge demands “epistemic iteration” (Chang 2004) through which tools are refined iteratively in tandem with the evolving comprehension of the phenomenon under study. Thus, with both optimism and caution, I look forward to future research that may refine the concept of bubbles in science and improve measurements.

Chapter 3

Chapter 3's interest shifts from biomedical science to the domain of machine learning research. Despite this transition, the chapter further explores the overarching theme of the dissertation in revealing the role of underlying structures in the contemporary scientific enterprise. The study conducted two strands of analysis. I first examine the extent to which the existence of code repositories, particularly GitHub repositories linked to papers, can enhance the monthly citation rates of focal research papers. This extends the previous literature studying the impact of data sharing on paper-level citations in the context of the open science movement. The second analysis is motivated by the observation that ML research increasingly relies on frameworks that allow the abstraction of technical details, thereby enabling researchers to deploy and experiment with their ideas more efficiently. Multiple open-source frameworks have been developed by various entities such as Google and Facebook (Meta) and compete with each other for so-called market shares. Drawing on the concept of second-order network effects (Katz and Shapiro 1986; Kauffman, McAndrews, and Wang 2000), this analysis investigates whether and to what extent the popularity of a particular ML framework boosts the citation rates of ML research papers.

With more straightforward research questions than the first and second chapters, the chapter aims to rigorously identify statistical patterns with a thorough linkage of the best available datasets: paper-repository records maintained by Papers with Code (PwC), Microsoft Academic Graph (MAG) and repository-level metadata collected from GitHub. The linkage process required substantial efforts and attention, as detailed in the chapter's appendix. The first analysis employs several statistical techniques: 1) random samples of papers accompanied with GitHub repositories, and 2) nearest neighbor matching for papers without repositories. The

machine particularly leverages vector representations of papers using the pre-trained embedding model applied to research topic keywords assigned by MAG, along with author affiliation information and team size. The results suggest an approximate 20% monthly citation advantage for papers with Github repositories compared to those without after the first repository became publicly available. The second analysis, evaluating the relationship between the popularity of ML frameworks and citation rates, reveals a modest unit magnitude of effect size that can lead to substantial (dis)advantages, depending on the framework used in code implementations. Overall, the findings underscore the significance of technological artifacts and infrastructure in research dissemination.

For this study, I initially planned to investigate the impact of a battery of repository-level characteristics—such as data folders, shell script files, inline codes, and code blocks within README files, alongside the inclusion of links to further resources and BibTeX references for easier citation—on academic paper citation rates. Working hypotheses included the dynamics of repository maintenance, authorship patterns, and their potential correlation with citation counts, for example, the effect of a paper's number of authors on repository quality, the influence of authors affiliated with tech companies on maintenance standards, and the overall relationship between repository maintenance quality and citation rates. However, it quickly became apparent that GitHub repositories manifest a substantial degree of uniformity, indicative of a broader trend toward standardization or isomorphism in documentation practices. This homogeneity reduces the variability required for meaningful analysis within the study's framework, and the final model approach uses fixed effects designed to absorb time-invariant characteristics present. As a result, despite the initial promise of these hypotheses to provide insights into the relationship between efforts for repository maintenance and citation trends, they were not directly tested.

Nonetheless, this uniformity has emerged as an interesting area for future research, signaling a broader shift towards standardized documentation practices that may merit detailed exploration.

A methodological consideration may challenge whether the study design accurately captures the "causal" effects of treatments— the presence of GitHub repositories and the popularity of ML frameworks—in a conventional manner. Ideally, field-level randomized controlled experiments would be employed. However, the feasibility of conducting such experiments or expecting two instances of a paper with the same context, title, and benchmarks—differing only in code availability—would be substantially limited. Nonetheless, I believe the study utilizes the best available data to quantify the impact of GitHub repositories and the association between the popularity of ML frameworks and citation rates, which has not been systematically studied before, contributing to the existing body of literature. It is crucial, however, to reemphasize here that the findings should not be interpreted as advocating for a monoculture in research, as discussed in the last section of the chapter.

REFERENCES

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2017. "The Dual Frontier: Patented Inventions and Prior Scientific Advance." *Science* 357 (6351): 583–87.

Anderson, Craig A., Akiko Shibuya, Nobuko Ihori, Edward L. Swing, Brad J. Bushman, Akira Sakamoto, Hannah R. Rothstein, and Muniba Saleem. 2010. "Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-Analytic Review." *Psychological Bulletin* 136 (2): 151–73.

Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*, edited by Orley C. Ashenfelter and David Card, 3:1277–1366. Elsevier.

Arroll, Bruce, C. Raina Elley, Tana Fishman, Felicity A. Goodyear-Smith, Tim Kenealy, Grant Blashki, Ngaire Kerse, and Stephen Macgillivray. 2009. "Antidepressants versus Placebo for Depression in Primary Care." *Cochrane Database of Systematic Reviews* , no. 3 (July): CD007954.

Arseniev-Koehler, Alina, and Jacob G. Foster. 2020. "Sociolinguistic Properties of Word Embeddings." *SocArXiv*. https://doi.org/10.31235/osf.io/b8kud.

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal* 99 (394): 116–31.

———. 1995. "Complexity in Economic and Financial Markets." *Complexity* 1 (1): 20–25.

Au, Larry. 2021. "Recent Scientific/intellectual Movements in Biomedicine." *Social Science & Medicine* 278 (June): 113950.

Azoulay, Pierre, Alessandro Bonatti, and Joshua L. Krieger. 2017. "The Career Effects of Scandal: Evidence from Scientific Retractions." *Research Policy* 46 (9): 1552–69.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *The American Economic Review* 109 (8): 2889–2920.

Azoulay, Pierre, Jeffrey L. Furman, and Fiona Murray. 2015. "Retractions." *The Review of Economics and Statistics* 97 (5): 1118–36.

Baker, Monya. 2016. "1,500 Scientists Lift the Lid on Reproducibility." *Nature* 533 (7604): 452–54.

Bates, Douglas, Deepayan Sarkar, Maintainer Douglas Bates, and L. Matrix. 2007. *The lme4 Package*.

Becker, Markus C., Nathalie Lazaric, Richard R. Nelson, and Sidney G. Winter. 2005. "Applying Organizational Routines in Understanding Organizational Change." *Industrial and Corporate Change* 14 (5): 775–91.

Belikov, Alexander V., Andrey Rzhetsky, and James Evans. 2022. "Prediction of Robust Scientific Facts from Literature." *Nature Machine Intelligence*, April, 1–10.

Belknap, Penny, and Wilbert M. Leonard. 1991. "A Conceptual Replication and Extension of Erving Goffman's Study of Gender Advertisements." *Sex Roles* 25 (3): 103–18.

Ben-David, Joseph, and Teresa A. Sullivan. 1975. "Sociology of Science." *Annual Review of Sociology* 1 (1): 203–22.

Bhattarai, Prajjwal, Mohammed Ghassemi, and Tuka Alhanai. 2022. "Open-Source Code Repository Attributes Predict Impact of Computer Science Research." In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, 1–7. JCDL '22. New York, NY, USA: Association for Computing Machinery.

Bishop, Bill. 2009. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us*

Apart. Houghton Mifflin Harcourt.
Bourdieu, Pierre. 1975. “The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason.” *Social Sciences Information. Information Sur Les Sciences Sociales* 14 (6): 19–47.
———. 1991. “The Peculiar History of Scientific Reason.” *Sociological Forum* 6 (1): 3–26.
———. 2004. *Science of Science and Reflexivity*. Polity.
Boutyline, Andrei, and Laura K. Soter. 2021. “Cultural Schemas: What They Are, How to Find Them, and What to Do Once You’ve Caught One.” *American Sociological Review* 86 (4): 728–58.
Breiger, Ronald L. 1974. “The Duality of Persons and Groups.” *Social Forces* 53 (2): 181–90.
Breiger, Ronald L., Scott A. Boorman, and Phipps Arabie. 1975. “An Algorithm for Clustering Relational Data with Applications to Social Network Analysis and Comparison with Multidimensional Scaling.” *Journal of Mathematical Psychology* 12 (3): 328–83.
Cai, Meng, Han Luo, Xiao Meng, Ying Cui, and Wei Wang. 2023. “Network Distribution and Sentiment Interaction: Information Diffusion Mechanisms between Social Bots and Human Users on Social Media.” *Information Processing & Management* 60 (2): 103197.
Cates, Christopher J., and Matthew J. Cates. 2008. “Regular Treatment with Salmeterol for Chronic Asthma: Serious Adverse Events.” *Cochrane Database of Systematic Reviews* , no. 3 (July): CD006363.
Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.
Chen, Xieling, Haoran Xie, Zongxi Li, Gary Cheng, Mingming Leng, and Fu Lee Wang. 2023. “Information Fusion and Artificial Intelligence for Smart Healthcare: A Bibliometric Study.” *Information Processing & Management* 60 (1): 103113.
Chien, Kenneth R., Jonas Frisén, Regina Fritsche-Danielson, Douglas A. Melton, Charles E. Murry, and Irving L. Weissman. 2019. “Regenerating the Field of Cardiovascular Cell Therapy.” *Nature Biotechnology* 37 (3): 232–37.
Choe, Yeongbae, Jooa Baek, and Hyesun Kim. 2023. “Heterogeneity in Consumer Preference toward Mega-Sport Event Travel Packages: Implications for Smart Tourism Marketing Strategy.” *Information Processing & Management* 60 (3): 103302.
Christensen, Garret, Allan Dafoe, Edward Miguel, Don A. Moore, and Andrew K. Rose. 2019. “A Study of the Impact of Data Sharing on Article Citations Using Journal Policies as a Natural Experiment.” *PloS One* 14 (12): e0225883.
Chu, Johan S. G., and James A. Evans. 2021. “Slowed Canonical Progress in Large Fields of Science.” *Proceedings of the National Academy of Sciences of the United States of America* 118 (41): e2021636118.
Cochrane, Archibald Leman. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.
Cochran, William G. 1954. “The Combination of Estimates from Different Experiments.” *Biometrics* 10 (1): 101–29.
Collins, Harry M. 1974. “The TEA Set: Tacit Knowledge and Scientific Networks.” *Science Studies* 4 (2): 165–85.
———. 1983. “The Sociology of Scientific Knowledge: Studies of Contemporary Science.” *Annual Review of Sociology* 9 (1): 265–85.
———. 1985. *Changing Order: Replication and Induction in Scientific Practice*. University of Chicago Press.

———. 2000. “Surviving Closure: Post-Rejection Adaptation and Plurality in Science.” *American Sociological Review* 65 (6): 824–45.
———. 2001. “Tacit Knowledge, Trust and the Q of Sapphire.” *Social Studies of Science* 31 (1): 71–85.
———. 2007. “Bicycling on the Moon: Collective Tacit Knowledge and Somatic-Limit Tacit Knowledge.” *Organization Studies* 28 (2): 257–62.
———. 2010. *Tacit and Explicit Knowledge*. University of Chicago Press.
Cosentino, Valerio, Javier L. Cánovas Izquierdo, and Jordi Cabot. 2017. “A Systematic Mapping Study of Software Development With GitHub.” *IEEE Access* 5: 7173–92.
Cowan, Robin, and Dominique Foray. 1997. “The Economics of Codification and the Diffusion of Knowledge.” *Industrial and Corporate Change* 6 (3): 595–622.
Crane, Diana. 1972. *Invisible Colleges; Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago Press.
Cui, Peng, Xiao Wang, Jian Pei, and Wenwu Zhu. 2019. “A Survey on Network Embedding.” *IEEE Transactions on Knowledge and Data Engineering* 31 (5): 833–52.
Daly, Jeanne. 2005. *Evidence-Based Medicine and the Search for a Science of Clinical Care*. University of California Press.
Danchev, Valentin, Andrey Rzhetsky, and James A. Evans. 2019. “Centralized Scientific Communities Are Less Likely to Generate Replicable Results.” *eLife* 8 (July): e43094.
Davidoff, F., B. Haynes, D. Sackett, and R. Smith. 1995. “Evidence Based Medicine.” *BMJ* 310 (6987): 1085–86.
David, Paul A. 1985. “Clio and the Economics of QWERTY.” *The American Economic Review* 75 (2): 332–37.
Davis, Fred D. 1989. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” *MIS Quarterly* 13 (3): 319–40.
Da, Zhi, and Xing Huang. 2020. “Harnessing the Wisdom of Crowds.” *Management Science* 66 (5): 1847–67.
Deeks, Jonathan J., and Julian P. T. Higgins. 2010. “Statistical Algorithms in Review Manager 5.” *Statistical Methods Group of The Cochrane Collaboration* 1 (11). https://training.cochrane.org/handbook/current/statistical-methods-revman5.
Díaz-Rodríguez, Natalia, Rūta Binkytė, Wafae Bakkali, Sannidhi Bookseller, Paola Tubaro, Andrius Bacevičius, Sami Zhioua, and Raja Chatila. 2023. “Gender and Sex Bias in COVID-19 Epidemiological Data through the Lenses of Causality.” *Information Processing & Management*, 103276.
Dickersin, K., R. Scherer, and C. Lefebvre. 1994. “Identifying Relevant Studies for Systematic Reviews.” *BMJ* 309 (6964): 1286–91.
Doing, Park. 2004. “‘lab Hands’ and the ‘Scarlet O.’” *Social Studies of Science* 34 (3): 299–323.
Dong, Xianlei, Jiahui Xu, Yi Bu, Chenwei Zhang, Ying Ding, Beibei Hu, and Yang Ding. 2022. “Beyond Correlation: Towards Matching Strategy for Causal Inference in Information Science.” *Journal of Information Science and Engineering* 48 (6): 735–48.
Dorch, Bertil F., Thea M. Drachen, and Ole Ellegaard. 2015. “The Data Sharing Advantage in Astrophysics.” *Proceedings of the International Astronomical Union* 11 (A29A): 172–75.
Economides, Nicholas, and Steven C. Salop. 1992. “Competition and Integration Among Complements, and Network Market Structure.” *The Journal of Industrial Economics* 40 (1): 105–23.
Eddy, D. M. 1990. “Practice Policies: Where Do They Come From?” *JAMA: The Journal of the*

American Medical Association 263 (9): 1265–75.
Evans, James P., Eric M. Meslin, Theresa M. Marteau, and Timothy Caulfield. 2011. "Genomics. Deflating the Genomic Bubble." *Science* 331 (6019): 861–62.
Evidence-Based Medicine Working Group. 1992. "Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine." *JAMA: The Journal of the American Medical Association* 268 (17): 2420–25.
Eyal, Gil. 2019. *The Crisis of Expertise*. John Wiley & Sons.
Fanelli, Daniele, Rodrigo Costas, and John P. A. Ioannidis. 2017. "Meta-Assessment of Bias in Science." *Proceedings of the National Academy of Sciences of the United States of America* 114 (14): 3714–19.
Fanelli, Daniele, and John P. A. Ioannidis. 2013. "US Studies May Overestimate Effect Sizes in Softer Research." *Proceedings of the National Academy of Sciences of the United States of America* 110 (37): 15031–36.
Färber, Michael. 2020. "Analyzing the GitHub Repositories of Research Papers." In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 491–92.
Feest, Uljana. 2016. "The Experimenters' Regress Reconsidered: Replication, Tacit Knowledge, and the Dynamics of Knowledge Generation." *Studies in History and Philosophy of Science* 58 (August): 34–45.
Fisher, Ronald Aylmer. 1925. *Statistical Methods for Research Workers*. New York, NY: Oliver & Boyd, Edinburgh & London.
———. 1935. *The Design of Experiments*. Oliver & Boyd, Edinburgh & London.
Fonseca Cacho, Jorge Ramón, and Kazem Taghva. 2018. "Reproducible Research in Document Analysis and Recognition." In *Information Technology - New Generations*, 389–95. Springer International Publishing.
Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, et al. 2018. "Science of Science." *Science* 359 (6379): eaao0185.
Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80 (5): 875–908.
Franklin, Allan. 1989. "The Epistemology of Experiment." In *The Uses of Experiment: Studies in the Natural Sciences*, edited by David Gooding, Trevor Pinch, and Simon Schaffer, 437–60. Cambridge University Press.
Franklin, Allan, and Harry M. Collins. 2016. "Two Kinds of Case Study and a New Agreement." In *The Philosophy of Historical Case Studies*, edited by Tilman Sauer and Raphael Scholl, 95–121. Cham: Springer International Publishing.
Frank, Morgan R., Dashun Wang, Manuel Cebrian, and Iyad Rahwan. 2019. "The Evolution of Citation Graphs in Artificial Intelligence Research." *Nature Machine Intelligence* 1 (2): 79–85.
Freeman, G. J., A. J. Long, Y. Iwai, K. Bourque, T. Chernova, H. Nishimura, L. J. Fitz, et al. 2000. "Engagement of the PD-1 Immunoinhibitory Receptor by a Novel B7 Family Member Leads to Negative Regulation of Lymphocyte Activation." *The Journal of Experimental Medicine* 192 (7): 1027–34.
Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70 (2): 204–32.
Fujimura, Joan H. 1988. "The Molecular Biological Bandwagon in Cancer Research: Where Social Worlds Meet." *Social Problems* 35 (3): 261–83.

Funk, Russell J., and Jason Owen-Smith. 2016. “A Dynamic Network Measure of Technological Change.” *Management Science* 63 (3): 791–817.
Furman, Jeffrey L., and Scott Stern. 2011. “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research.” *The American Economic Review* 101 (5): 1933–63.
Galton, Francis. 1907. “Vox Populi (the Wisdom of Crowds).” *Nature* 75 (7): 450–51.
Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (16): E3635–44.
Glass, Gene V. 1976. “Primary, Secondary, and Meta-Analysis of Research.” *Educational Researcher* 5 (10): 3–8.
Goldacre, Ben. 2014. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*. Macmillan.
Goldman, Alvin I., and Moshe Shaked. 1991. “An Economic Model of Scientific Activity and Truth Acquisition.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 63 (1): 31–55.
Gonzalez, Danielle, Thomas Zimmermann, and Nachiappan Nagappan. 2020. “The State of the ML-Universe: 10 Years of Artificial Intelligence & Machine Learning Software Development on GitHub.” In *Proceedings of the 17th International Conference on Mining Software Repositories*, 431–42. MSR ’20. New York, NY, USA: Association for Computing Machinery.
Greenhalgh, Joanne, Rob Flynn, Andrew F. Long, and Sarah Tyson. 2008. “Tacit and Encoded Knowledge in the Use of Standardised Outcome Measures in Multidisciplinary Team Decision Making: A Case Study of in-Patient Neurorehabilitation.” *Social Science & Medicine* 67 (1): 183–94.
Green, H. N., D. Pindar, G. Davis, and E. Mellanby. 1931. “DIET AS A PROPHYLACTIC AGENT AGAINST PUERPERAL SEPSIS.” *British Medical Journal* 2 (3691): 595–98.
Grover, Aditya, and Jure Leskovec. 2016. “node2vec: Scalable Feature Learning for Networks.” *KDD: Proceedings / International Conference on Knowledge Discovery & Data Mining. International Conference on Knowledge Discovery & Data Mining* 2016 (August): 855–64.
Gurevitch, Jessica, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. 2018. “Meta-Analysis and the Science of Research Synthesis.” *Nature* 555 (7695): 175–82.
Hacking, Ian. 1988. “Telepathy: Origins of Randomization in Experimental Design.” *Isis; an International Review Devoted to the History of Science and Its Cultural Influences* 79 (3): 427–51.
Haibe-Kains, Benjamin, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors, Levi Waldron, Bo Wang, et al. 2020. “Transparency and Reproducibility in Artificial Intelligence.” *Nature* 586 (7829): E14–16.
Hall, Nancy S. 2007. “R. A. Fisher and His Advocacy of Randomization.” *Journal of the History of Biology* 40 (2): 295–325.
Hann, Michael M. 2011. “Molecular Obesity, Potency and Other Addictions in Drug Discovery.” *MedChemComm* 2 (5): 349–55.
Harras, Georges, and Didier Sornette. 2011. “How to Grow a Bubble: A Model of Myopic Adapting Agents.” *Journal of Economic Behavior & Organization* 80 (1): 137–52.
Harris, Richard. 2017. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes*

Hope, and Wastes Billions. Hachette UK.
Hausman, Jerry, Bronwyn Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." National Bureau of Economic Research. https://doi.org/10.3386/t0017.
Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. 2015. "The Extent and Consequences of P-Hacking in Science." *PLoS Biology* 13 (3): e1002106.
Heesen, Remco. 2018. "Why the Reward Structure of Science Makes Reproducibility Problems Inevitable." *The Journal of Philosophy* 115 (12): 661–74.
He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Deep Residual Learning for Image Recognition." *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1512.03385.
Henneken, Edwin A., and Alberto Accomazzi. 2011. "Linking to Data - Effect on Citation Rates in Astronomy." *arXiv [cs.DL]*. arXiv. http://arxiv.org/abs/1111.3618.
Higgins, Julian P. T., James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, and Vivian A. Welch. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons.
Higgins, Julian P. T., and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21 (11): 1539–58.
Higgins, Julian P. T., Simon G. Thompson, Jonathan J. Deeks, and Douglas G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *BMJ* 327 (7414): 557–60.
Hilgard, Joseph, Christopher R. Engelhardt, and Jeffrey N. Rouder. 2017. "Overstated Evidence for Short-Term Effects of Violent Games on Affect and Behavior: A Reanalysis of Anderson et Al. (2010)." *Psychological Bulletin* 143 (7): 757–74.
Hill, A. Bradford. 1952. "The Clinical Trial." *The New England Journal of Medicine* 247 (4): 113–19.
Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97 (460): 1090–98.
Huang, Shengzhi, Wei Lu, Yi Bu, and Yong Huang. 2022. "Revisiting the Exploration-Exploitation Behavior of Scholars' Research Topic Selection: Evidence from a Large-Scale Bibliographic Database." *Information Processing & Management* 59 (6): 103110.
Huang, Yong, Wei Lu, Jialin Liu, Qikai Cheng, and Yi Bu. 2022. "Towards Transdisciplinary Impact of Scientific Publications: A Longitudinal, Comprehensive, and Large-Scale Analysis on Microsoft Academic Graph." *Information Processing & Management* 59 (2): 102859.
Hughes, Peyton, Damian Marshall, Yvonne Reid, Helen Parkes, and Cohava Gelber. 2007. "The Costs of Using Unauthenticated, over-Passaged Cell Lines: How Much More Data Do We Need?" *BioTechniques* 43 (5): 575–86.
Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis*. Russell Sage Foundation.
Hutchins, B. Ian, Matthew T. Davis, Rebecca A. Meseroll, and George M. Santangelo. 2019. "Predicting Translational Progress in Biomedical Research." *PLoS Biology* 17 (10): e3000416.
Hutchins, B. Ian, Xin Yuan, James M. Anderson, and George M. Santangelo. 2016. "Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level." *PLoS Biology* 14 (9): e1002541.
Hutson, Matthew. 2018. "Artificial Intelligence Faces Reproducibility Crisis." *Science* 359

(6377): 725–26.
Ioannidis, John P. A. 2005. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2 (8): e124.
Iwai, Yoshiko, Junzo Hamanishi, Kenji Chamoto, and Tasuku Honjo. 2017. “Cancer Immunotherapies Targeting the PD-1 Signaling Pathway.” *Journal of Biomedical Science* 24 (1): 1–11.
Jepson, Ruth G., Gabrielle Williams, and Jonathan C. Craig. 2012. “Cranberries for Preventing Urinary Tract Infections.” *Cochrane Database of Systematic Reviews* 10 (10): CD001321.
Jiang, Qiaolei, Yadi Zhang, and Wenjing Pian. 2022. “Chatbot as an Emergency Exist: Mediated Empathy for Resilience via Human-AI Interaction during the COVID-19 Pandemic.” *Information Processing & Management* 59 (6): 103074.
Jones, Benjamin F., and Lawrence H. Summers. 2020. “A Calculation of the Social Returns to Innovation.” National Bureau of Economic Research. https://doi.org/10.3386/w27863.
Junod, Suzanne White. 2008. “FDA and Clinical Drug Trials: A Short History.” *A Quick Guide to Clinical Trials*, 25–55.
Kang, Donghyun, Taeyoung Kang, and Junkyu Jang. 2023. “Papers with Code or without Code? Impact of GitHub Repository Usability on the Diffusion of Machine Learning Research.” *Information Processing & Management* 60 (6): 103477. https://doi.org/10.1016/j.ipm.2023.103477
Kantrowitz, A. 1967. “Proposal for an Institution for Scientific Judgment.” *Science* 156 (3776): 763–64.
Kapoor, Sayash, and Arvind Narayanan. 2022. “Leakage and the Reproducibility Crisis in ML-Based Science.” *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2207.07048.
Katz, Michael L., and Carl Shapiro. 1985. “Network Externalities, Competition, and Compatibility.” *The American Economic Review* 75 (3): 424–40.
———. 1986. “Technology Adoption in the Presence of Network Externalities.” *Journal of Political Economy* 94 (4): 822–41.
Kauffman, Robert J., James McAndrews, and Yu-Ming Wang. 2000. “Opening the ‘Black Box’ of Network Externalities in Network Adoption.” *Information Systems Research* 11 (1): 61–82.
Kepes, Sven, Brad J. Bushman, and Craig A. Anderson. 2017. “Violent Video Game Effects Remain a Societal Concern: Reply to Hilgard, Engelhardt, and Rouder (2017).” *Psychological Bulletin* 143 (7): 775–82.
Kim, Hee-Woong, Hock Chuan Chan, and Sumeet Gupta. 2007. “Value-Based Adoption of Mobile Internet: An Empirical Investigation.” *Decision Support Systems* 43 (1): 111–26.
Kim, Junsol, Zhao Wang, Haohan Shi, Hsin-Keng Ling, and James Evans. 2023. “Individual Misinformation Tagging Reinforces Echo Chambers; Collective Tagging Does Not.” *arXiv [cs.CY]*. arXiv. http://arxiv.org/abs/2311.11282.
Kim, Kyung-Man. 1994. *Explaining Scientific Consensus: The Case of Mendelian Genetics*. Guilford Publications.
Kim, Youngseek, and Melissa Adler. 2015. “Social Scientists’ Data Sharing Behaviors: Investigating the Roles of Individual Motivations, Institutional Pressures, and Data Repositories.” *International Journal of Information Management* 35 (4): 408–18.
Klavans, Richard, Kevin W. Boyack, and Dewey A. Murdick. 2020. “A Novel Approach to Predicting Exceptional Growth in Research.” *PloS One* 15 (9): e0239177.
Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and*

Contextual Nature of Science. Elsevier Science & Technology Books.
———. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.
Koch, Bernard, Emily Denton, Alex Hanna, and Jacob G. Foster. 2021. "Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research." *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2112.01716.
Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84 (5): 905–49.
Krauss, Alexander. 2018. "Why All Randomised Controlled Trials Produce Biased Results." *Annals of Medicine* 50 (4): 312–22.
Krogh, Georg von, Stefan Haefliger, Sebastian Spaeth, and Martin W. Wallin. 2012. "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development." *MIS Quarterly* 36 (2): 649–76.
Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press.
Kwon, Seokbeom, and Kazuyuki Motohashi. 2021. "Incentive or Disincentive for Research Data Disclosure? A Large-Scale Empirical Analysis and Implications for Open Science Policy." *International Journal of Information Management* 60 (October): 102371.
Latour, Bruno, and Steven Woolgar. 1979. *Laboratory Life: The Social Construction of Social Facts*. Sage.
Lazarsfeld, Paul F., and Neil W. Henry. 1968. *Latent Structure Analysis*. Houghton Mifflin.
Le, Quoc, and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." In *Proceedings of the 31st International Conference on Machine Learning*, edited by Eric P. Xing and Tony Jebara, 32:1188–96. Proceedings of Machine Learning Research. Bejing, China: PMLR.
Levy, Omer, and Yoav Goldberg. 2014. "Neural Word Embedding as Implicit Matrix Factorization." In *Advances in Neural Information Processing Systems*. Vol. 27. https://proceedings.neurips.cc/paper/2014/hash/feab05aa91085b7a8012516bc3533958-Abstract.html.
Lin, Jimmy, and W. John Wilbur. 2007. "PubMed Related Articles: A Probabilistic Topic-Based Model for Content Similarity." *BMC Bioinformatics* 8 (October): 1–14.
Lin, Yiling, James A. Evans, and Lingfei Wu. 2022. "New Directions in Science Emerge from Disconnection and Discord." *Journal of Informetrics* 16 (1): 101234.
Li, Xin, Traci J. Hess, and Joseph S. Valacich. 2008. "Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems." *The Journal of Strategic Information Systems* 17 (1): 39–71.
Lix, Katharina, Amir Goldberg, Sameer B. Srivastava, and Melissa A. Valentine. 2022. "Aligning Differences: Discursive Diversity and Team Performance." *Management Science*, February, 8430–48.
Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences of the United States of America* 108 (22): 9020–25.
MacKenzie, Donald, and Graham Spinardi. 1995. "Tacit Knowledge, Weapons Design, and the Uninvention of Nuclear Weapons." *The American Journal of Sociology* 101 (1): 44–99.
Mannheim, Karl. (1936) 1991. *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. Translated by Wirth Louis and Shils Edward. London, England and New York:

Routledge.
Martínez-Plumed, Fernando, Pablo Barredo, Seán Ó. hÉigeartaigh, and José Hernández-Orallo. 2021. “Research Community Dynamics behind Popular AI Benchmarks.” *Nature Machine Intelligence* 3 (7): 581–89.
Martin, John Levi. 2003. “What Is Field Theory?” *The American Journal of Sociology* 109 (1): 1–49.
McIntyre, David P., and Arati Srinivasan. 2017. “Networks, Platforms, and Strategy: Emerging Views and next Steps.” *Strategic Management Journal* 38 (1): 141–60.
McKiernan, Erin C., Philip E. Bourne, C. Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, et al. 2016. “How Open Science Helps Researchers Succeed.” *eLife* 5 (July): e16800.
McMahan, Peter, and Daniel A. McFarland. 2021. “Creative Destruction: The Structural Consequences of Scientific Curation.” *American Sociological Review* 86 (2): 341–76.
Meldrum, Marcia. 1998. “‘A Calculated Risk’: The Salk Polio Vaccine Field Trials of 1954.” *BMJ* 317 (7167): 1233–36.
Menchik, Daniel A. 2021. *Managing Medical Authority: How Doctors Compete for Status and Create Knowledge*. Princeton, NJ: Princeton University Press.
Merton, Robert K. 1957. “Priorities in Scientific Discovery: A Chapter in the Sociology of Science.” *American Sociological Review* 22 (6): 635–59.
———. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.
Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.” *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1301.3781.
Min, Chao, Yi Bu, Ding Wu, Ying Ding, and Yi Zhang. 2021. “Identifying Citation Patterns of Scientific Breakthroughs: A Perspective of Dynamic Citation Process.” *Information Processing & Management* 58 (1): 102428.
Mitroff, Ian I. 1974. “Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists.” *American Sociological Review* 39 (4): 579–95.
Montgomery, Kathryn. 2006. *How Doctors Think: Clinical Judgment and the Practice of Medicine*. OUP USA.
Mueller-Langer, Frank, Benedikt Fecher, Dietmar Harhoff, and Gert G. Wagner. 2019. “Replication Studies in economics—How Many and Which Papers Are Chosen for Replication, and Why?” *Research Policy* 48 (1): 62–83.
Mukherjee, Arijit, and Scott Stern. 2009. “Disclosure or Secrecy? The Dynamics of Open Science.” *International Journal of Industrial Organization* 27 (3): 449–62.
Mulrow, C. D. 1994. “Rationale for Systematic Reviews.” *BMJ* 309 (6954): 597–99.
Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. “A Manifesto for Reproducible Science.” *Nature Human Behaviour* 1 (1): 1–9.
Murry, Charles E., Mark H. Soonpaa, Hans Reinecke, Hidehiro Nakajima, Hisako O. Nakajima, Michael Rubart, Kishore B. S. Pasumarthi, et al. 2004. “Haematopoietic Stem Cells Do Not Transdifferentiate into Cardiac Myocytes in Myocardial Infarcts.” *Nature* 428 (6983): 664–68.

Myers, Kyle. 2020. “The Elasticity of Science.” *American Economic Journal. Applied Economics* 12 (4): 103–34.
Neimark, Jill. 2015. “Line of Attack.” *Science* 347 (6225): 938–40.
Nelson, Laura K. 2021. “Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South.” *Poetics* 88 (October): 101539.
Nelson, Richard R. 2003. “On the Uneven Evolution of Human Know-How.” *Research Policy* 32 (6): 909–22.
Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716.
Oreskes, Naomi. 2019. *Why Trust Science?* Princeton University Press.
Orlic, D., J. Kajstura, S. Chimenti, I. Jakoniuk, S. M. Anderson, B. Li, J. Pickel, et al. 2001. “Bone Marrow Cells Regenerate Infarcted Myocardium.” *Nature* 410 (6829): 701–5.
Osafune, Kenji, Leslie Caron, Malgorzata Borowiak, Rita J. Martinez, Claire S. Fitz-Gerald, Yasunori Sato, Chad A. Cowan, Kenneth R. Chien, and Douglas A. Melton. 2008. “Marked Differences in Differentiation Propensity among Human Embryonic Stem Cell Lines.” *Nature Biotechnology* 26 (3): 313–15.
Pachucki, Mark A., and Ronald L. Breiger. 2010. “Cultural Holes: Beyond Relationality in Social Networks and Culture.” *Annual Review of Sociology* 36 (1): 205–24.
Page, Scott E. 2019. *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy*. Princeton University Press.
Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.
Partha, Dasgupta, and Paul A. David. 1994. “Toward a New Economics of Science.” *Research Policy* 23 (5): 487–521.
Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. “Data and Its (dis)contents: A Survey of Dataset Development and Use in Machine Learning Research.” *Patterns* 2 (11): 100336.
Pautasso, Marco. 2010. “Worsening File-Drawer Problem in the Abstracts of Natural, Medical and Social Science Databases.” *Scientometrics* 85 (1): 193–202.
Pavitt, Keith. 1987. “The Objectives of Technology Policy.” *Science & Public Policy* 14 (4): 182–88.
Pedersen, David Budtz, and Vincent F. Hendricks. 2014. “Science Bubbles.” *Philosophy & Technology* 27 (4): 503–18.
Peirce, Charles Sanders, and Joseph Jastrow. 1884. “On Small Differences in Sensation.” *Memoirs of the National Academy of Sciences* 3. https://philarchive.org/archive/PEIOSD.
Peng, Gang. 2019. “Co-Membership, Networks Ties, and Knowledge Flow: An Empirical Investigation Controlling for Alternative Mechanisms.” *Decision Support Systems* 118 (March): 83–90.
Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. “Glove: Global Vectors for Word Representation.” In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–43.
Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. 2014. “DeepWalk: Online Learning of Social Representations.” In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–10. KDD ’14. New York, NY, USA: Association for Computing Machinery.
Petersen, Alexander M., Raj K. Pan, Fabio Pammolli, and Santo Fortunato. 2019. “Methods to

Account for Citation Inflation in Research Evaluation." *Research Policy* 48 (7): 1855–65.
Peterson, David, and Aaron Panofsky. 2021. "Self-Correction in Science: The Diagnostic and Integrative Motives for Replication." *Social Studies of Science*, March, 583–605.
Pineau, Joelle, Philippe Vincent-Lamarre, Koustuv Sinha, Vincent Larivière, Alina Beygelzimer, Florence d'Alché-Buc, Emily Fox, and Hugo Larochelle. 2021. "Improving Reproducibility in Machine Learning Research: A Report from the NeurIPS 2019 Reproducibility Program." *Journal of Machine Engineering* 22 (1): 7459–78.
Piwowar, Heather A., and Todd J. Vision. 2013. "Data Reuse and the Open Data Citation Advantage." *PeerJ* 1 (October): e175.
Polanyi, Michael. 1958. *Personal Knowledge*. Routledge.
Polanyi, Michael, and Amartya Sen. 1966. *The Tacit Dimension*. University of Chicago Press.
Quaini, Federico, Konrad Urbanek, Antonio P. Beltrami, Nicoletta Finato, Carlo A. Beltrami, Bernardo Nadal-Ginard, Jan Kajstura, Annarosa Leri, and Piero Anversa. 2002. "Chimerism of the Transplanted Heart." *New England Journal of Medicine* 346 (1): 5–15.
Radim Rehurek, Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In *PROCEEDINGS OF THE LREC 2010 WORKSHOP ON NEW CHALLENGES FOR NLP FRAMEWORKS*.
Ragu-Nathan, T. S., Monideepa Tarafdar, Bhanu S. Ragu-Nathan, and Qiang Tu. 2008. "The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation." *Information Systems Research* 19 (4): 417–33.
Reschke, Brian P., Pierre Azoulay, and Toby E. Stuart. 2018. "Status Spillovers: The Effect of Status-Conferring Prizes on the Allocation of Attention." *Administrative Science Quarterly* 63 (4): 819–47.
Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638–41.
Rzhetsky, Andrey, Ivan Iossifov, Ji Meng Loh, and Kevin P. White. 2006. "Microparadigms: Chains of Collective Reasoning in Publications about Molecular Interactions." *Proceedings of the National Academy of Sciences of the United States of America* 103 (13): 4940–45.
Sackett, D. L., W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson. 1996. "Evidence Based Medicine: What It Is and What It Isn't." *BMJ* 312 (7023): 71–72.
Salandra, Rossella, Paola Criscuolo, and Ammon Salter. 2021. "Directing Scientists Away from Potentially Biased Publications: The Role of Systematic Reviews in Health Care." *Research Policy* 50 (1): 104130.
Shadish, William R., and Jesse D. Lecy. 2015. "The Meta-Analytic Big Bang." *Research Synthesis Methods* 6 (3): 246–64.
Shenhav, Liat, Ruth Heller, and Yoav Benjamini. 2015. "Quantifying Replicability in Systematic Reviews: The R-Value." *arXiv [stat.AP]*. arXiv. http://arxiv.org/abs/1502.00088.
Shen, Zhihong, Hao Ma, and Kuansan Wang. 2018. "A Web-Scale System for Scientific Knowledge Exploration." *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1805.12216.
Shwed, Uri, and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review* 75 (6): 817–40.
Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (paul) Hsu, and Kuansan Wang. 2015. "An Overview of Microsoft Academic Service (MAS) and Applications." In *Proceedings of the 24th International Conference on World Wide Web*, 243–46.
Smaill, Fiona M., and Rosalie M. Grivell. 2014. "Antibiotic Prophylaxis versus No Prophylaxis

for Preventing Infection after Cesarean Section." *Cochrane Database of Systematic Reviews* 2014 (10): CD007482.
Small, Henry, Kevin W. Boyack, and Richard Klavans. 2014. "Identifying Emerging Topics in Science and Technology." *Research Policy* 43 (8): 1450–67.
Smith, Anna L., Dena M. Asta, and Catherine A. Calder. 2019. "The Geometry of Continuous Latent Space Models for Network Data." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 34 (3): 428–53.
Solla Price, Derek John de. 1961. *Science Since Babylon*. New Haven and London : Yale University Press.
Solla Price, Derek John de, and Donald Beaver. 1966. "Collaboration in an Invisible College." *The American Psychologist* 21 (11): 1011–18.
Solomon, Miriam. 2015. *Making Medical Knowledge*. Oxford University Press.
Soonpaa, Mark H., Michael Rubart, and Loren J. Field. 2013. "Challenges Measuring Cardiomyocyte Renewal." *Biochimica et Biophysica Acta* 1833 (4): 799–803.
Stodden, Victoria, Marcia McNutt, David H. Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A. Heroux, John P. A. Ioannidis, and Michela Taufer. 2016. "Enhancing Reproducibility for Computational Methods." *Science* 354 (6317): 1240–41.
Stoughton, Jason. 2022. "The Bubble-Bursting, Causality-Revealing Awesomeness of Randomized Controlled Trials." National Science Foundation. https://new.nsf.gov/science-matters/bubble-bursting-causality-revealing-awesomeness.
Sunstein, Cass R. 2018. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business*. New York: Doubleday.
Taylor, Marisa, and Brad Heath. 2022. "Years after Brigham-Harvard Scandal, U.S. Pours Millions into Tainted Stem-Cell Field." *Reuters*, June 21, 2022. https://www.reuters.com/investigates/special-report/health-hearts-stem-cells/.
Tennant, Jonathan P., Jonathan M. Dugan, Daniel Graziotin, Damien C. Jacques, François Waldner, Daniel Mietchen, Yehia Elkhatib, et al. 2017. "A Multi-Disciplinary Perspective on Emergent and Future Innovations in Peer Review." *F1000Research* 6 (July): 1151.
Teplitskiy, Misha, Daniel Acuna, Aïda Elamrani-Raoult, Konrad Körding, and James Evans. 2018. "The Sociology of Scientific Validity: How Professional Networks Shape Judgement in Peer Review." *Research Policy* 47 (9): 1825–41.
The Cochrane Collaboration. 2020. *Review Manager (RevMan) [Computer Program]. Version 5.4*.
Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge Of Evidence-Based Medicine*. Temple University Press.
Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty." *Science* 185 (4157): 1124–31.
Vandewalle, Patrick. 2012. "Code Sharing Is Associated with Research Impact in Image Processing." *Computing in Science & Engineering* 14 (4): 42–47.
———. 2019. "Code Availability for Image Processing Papers: A Status Update." *WIC IEEE SP Symposium on Information Theory*. https://lirias.kuleuven.be/2815281?limo=0.
Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. "User Acceptance of Information Technology: Toward a Unified View." *MIS Quarterly* 27 (3):

425–78.
Viechtbauer, Wolfgang. 2010. “Conducting Meta-Analyses in R with the Metafor Package.” *Journal of Statistical Software* 36 (3): 1–48.
Vrieze, J. 2018. “Meta-Analyses Were Supposed to End Scientific Debates. Often, They Only Cause More Controversy.” *Science*, September 18, 2018.
Wang, Dashun, and Albert-László Barabási. 2021. *The Science of Science*. Cambridge University Press.
Wang, Kuansan, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Darrin Eide, Yuxiao Dong, Junjie Qian, Anshul Kanakia, Alvin Chen, and Richard Rogahn. 2019. “A Review of Microsoft Academic Services for Science of Science Studies.” *Frontiers in Big Data* 2 (December): 45.
Wang, Wei, Lihuan Guo, Yenchun Jim Wu, Mark Goh, and Shouyi Wang. 2022. “Content-Oriented or Persona-Oriented? A Text Analytics of Endorsement Strategies on Public Willingness to Participate in Citizen Science.” *Information Processing & Management* 59 (2): 102832.
Warwick, Andrew. 2011. *Masters of Theory: Cambridge and the Rise of Mathematical Physics*. University of Chicago Press.
Weis, James W., and Joseph M. Jacobson. 2021. “Learning on Knowledge Graph Dynamics Provides an Early Warning of Impactful Research.” *Nature Biotechnology* 39 (10): 1300–1307.
Wennberg, J. E. 1984. “Dealing with Medical Practice Variations: A Proposal for Action.” *Health Affairs* 3 (2): 6–32.
West, Suzanne L., Gerald Gartlehner, Alyssa J. Mansfield, Charles Poole, Elizabeth Tant, Nancy Lenfestey, Linda J. Lux, et al. 2011. *Comparative Effectiveness Review Methods: Clinical Heterogeneity*. Rockville (MD): Agency for Healthcare Research and Quality (US).
Wilcken, N., J. Hornbuckle, and D. Ghersi. 2003. “Chemotherapy Alone versus Endocrine Therapy Alone for Metastatic Breast Cancer.” *Cochrane Database of Systematic Reviews* , no. 2: CD002747.
Wilms, Konstantin L., Stefan Stieglitz, Björn Ross, and Christian Meske. 2020. “A Value-Based Perspective on Supporting and Hindering Factors for Research Data Management.” *International Journal of Information Management* 54 (October): 102174.
Worrall, John. 2002. “What Evidence in Evidence-Based Medicine?” *Philosophy of Science* 69 (S3): S316–30.
Xie, Zilin, Wenguo Weng, Yufeng Pan, Zhiyuan Du, Xingyi Li, and Yijian Duan. 2023. “Public Opinion Changing Patterns under the Double-Hazard Scenario of Natural Disaster and Public Health Event.” *Information Processing & Management* 60 (3): 103287.
Xu, Jian, Sunkyu Kim, Min Song, Minbyul Jeong, Donghyeon Kim, Jaewoo Kang, Justin F. Rousseau, et al. 2020. “Building a PubMed Knowledge Graph.” *Scientific Data* 7 (1): 205.
Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. “Recent Trends in Deep Learning Based Natural Language Processing.” *IEEE Computational Intelligence Magazine* 13 (3): 55–75.
Zhang, Chenwei, Yi Bu, Ying Ding, and Jian Xu. 2018. “Understanding Scientific Collaboration: Homophily, Transitivity, and Preferential Attachment.” *Journal of the Association for Information Science and Technology* 69 (1): 72–86.
Zhang, Xuanhui, Weijia Zhang, Yuxiang (chris) Zhao, and Qinghua Zhu. 2022. “Imbalanced Volunteer Engagement in Cultural Heritage Crowdsourcing: A Task-Related Exploration

Based on Causal Inference." *Information Processing & Management* 59 (5): 103027.
Zhao, Zhenyue, Yi Bu, and Jiang Li. 2021. "Characterizing Scientists Leaving Science before Their Time: Evidence from Mathematics." *Information Processing & Management* 58 (5): 102661.