The Decline in the Transmission of Scientific Ideas*

Enrico Berkes† Ruben Gaetani‡

March 2026

Abstract

We document that the diffusion of new scientific ideas beyond their field of origin has declined substantially over the past four decades. This contraction is closely linked to increasing specialization in scientific language: research that employs more technical terminology tends to be adopted less broadly. We develop a theory of scientific discovery in which the diffusion of new ideas depends on the degree to which potential adopters can understand and process them. When introducing their discoveries, scientists face a tradeoff between technical communication targeted at their immediate peers and more accessible language meant to reach broader audiences. As knowledge accumulates and research at the frontier builds on deeper layers of prior work, this tradeoff increasingly favors specialized language, limiting diffusion. Policy interventions that align scientists' incentives can broaden adoption and increase the social value of scientific research.

JEL Codes: O3, O4

Keywords: Scientific Ideas, Idea Transmission, Scientific Language

*We thank for their comments Ben Jones, Scott Stern, Aakash Kalyani, Michael Andrews and seminar participants at the University of Toronto, Brown University, Ohio State University, University of California Merced, Università della Svizzera Italiana, Towson University, Boise State University, Innovation Workshop (University of Luxembourg), and 20th Conference of Swiss Economists Abroad. Jayant Yadav provided outstanding research assistance. Ruben Gaetani is supported in part by funding from the Social Sciences and Humanities Research Council of Canada.

†University of Maryland, Baltimore County, enrico.berkes@umbc.edu.

‡University of Toronto, ruben.gaetani@utoronto.ca.

1 Introduction

Many influential scientific discoveries generated their largest benefits by finding application outside the boundaries of their original field. X-ray crystallography, developed in physics to study the atomic structure of crystals, became a central tool in molecular biology and made it possible to infer the double-helix structure of DNA. The fast Fourier transform, introduced as a computational tool in electrical engineering, entered medical diagnostics by enabling the conversion of magnetic signals from the body into images. More recently, CRISPR, first identified as part of basic research in bacterial immune systems, has become a widely used tool for targeted genetic editing in many areas of biology, medicine, and agriculture.

While much of the social value of science stems from ideas diffusing beyond their originally intended scope, the forces that govern this process remain poorly understood. Existing work has documented a range of barriers to adoption, such as constraints to accessing data and scientific knowledge (Furman and Stern, 2011; Nagaraj et al., 2020; Berkes and Nencka, 2024), institutional boundaries that discourage pivoting (Myers, 2020; Hill et al., 2025), and limited opportunities for knowledge exchange across disciplines (Catalini, 2018; Lane et al., 2021; Berkes and Gaetani, 2021). Yet we have limited evidence on what gives rise to these barriers, how they change over time, and how patterns of diffusion evolve as a result. These questions have become increasingly salient as the returns to innovation appear to have declined in recent decades, a trend that has been linked to the observed slowdown in productivity growth (Jones, 2009; Bloom et al., 2020; Ekerdt and Wu, 2024; Fort et al., 2025).

In this paper, we study the diffusion of science starting from a simple premise: the breadth of adoption of new ideas depends on how intelligible they are to audiences outside their originating field. We develop a theory of scientific discovery in which *comprehension*—the extent to which potential adopters can understand and process a new idea in a way that enables application—depends on both the depth of the accumulated knowledge behind the idea and the accessibility of the language used to introduce it. As the stock of knowledge increases and new ideas are built on deeper layers of prior work, scientists optimally favor more specialized language, narrowing diffusion. Using data on scientific publications from PubMed, we find strong support for this mechanism. Over the last four decades, the breadth of adoption of new scientific ideas has declined substantially, and changes in language accessibility have been an important contributing factor to this decline. These findings imply that the social returns to science may fall even when the intrinsic

value of new ideas remains unchanged, opening room for policy to align scientists' incentives with the social benefits of broad diffusion.

That scientists make deliberate choices about how to communicate their research, and that these choices are consequential for the impact of their work, is an observation that recurs throughout the history of science. The philosopher of science Thomas Kuhn argued that the adoption of a new paradigm requires shared frameworks and persuasive arguments that facilitate mutual understanding (Kuhn, 1962). Richard Feynman claimed that his introduction of diagrammatic methods was essential for making quantum electrodynamics usable beyond a small group of specialists, allowing physicists in other areas to apply and extend the theory.[1] By contrast, the work of Josiah Willard Gibbs on statistical mechanics was characterized by such mathematical density that even Albert Einstein famously found it "inscrutable," contributing to a long delay before the scientific community could fully integrate his insights.

Why do different ideas vary so markedly in how they are communicated? Our central argument is that these differences reflect the tradeoffs scientists face when conveying their discoveries. The goal of our theory is to study these tradeoffs and their implications for the diffusion of science.

In the model, a new idea is developed with potential applications across multiple fields. Adoption occurs when potential users achieve a level of comprehension sufficient to apply the idea in their own domain. Comprehension depends on three elements: the *scientific distance* between the field of origin and the field of potential adoption, the idea's *knowledge depth*—that is, the degree to which the idea is detached from basic and self-evident principles—and the *accessibility* of the language used to communicate it.

Comprehension declines with scientific distance, as ideas become harder to understand when the cognitive frameworks of potential adopters differ from that of the originating field. This decline is steeper for ideas with greater knowledge depth, because such ideas rely more heavily on the assumptions and methods embedded in the originating field's cognitive framework. Communication choices interact with these forces. By using more accessible language, scientists can bridge cognitive distance and make ideas more comprehensible to distant audiences (*bridging effect*), at the cost of reduced precision and technical detail, which lowers comprehension for all audiences (*dilution effect*). On net, more accessible language increases comprehension for distant audiences but lowers it for nearby ones. Both the benefits of bridging and the costs of dilution are amplified as the

[1] Frank Wilczek noted that key calculations underlying his Nobel Prize-winning work would have been "unthinkable without Feynman diagrams" (Frank Wilczek, "How Feynman Diagrams Almost Saved Space", Quanta Magazine, 2016, accessed January 2026).

depth of knowledge underlying the idea increases.

From this setting, we derive predictions on the joint evolution of knowledge depth, communication choices, and diffusion. First, when the dilution effect is more sensitive to knowledge depth than the bridging effect, scientists will optimally choose to communicate ideas built on deeper layers of knowledge using less accessible language. Intuitively, the loss of precision associated with simplifying language becomes more severe as ideas grow more detached from self-evident principles, prompting scientists to resolve the communication tradeoff in favor of technical language. This response is exacerbated by scientists' own-field bias—the degree to which adoption by distant fields is discounted relative to adoption by nearby ones. Second, more accessible language expands the breadth of adoption by making ideas more comprehensible to distant audiences. The combination of these effects implies that rising knowledge depth reduces the breadth of adoption and, if dilution costs are sufficiently responsive, changes in communication operate as a mediating channel.

We bring these insights to the data using a corpus of scientific publications indexed in PubMed. We identify newly introduced scientific concepts and track their adoption across fields and over time. For each concept, we construct measures that correspond to the objects formalized in the model. Knowledge depth—which is difficult to observe directly—is proxied by the degree of specialization of the authors introducing the concept. Language accessibility is measured using the prevalence of common words in the abstracts of the papers that introduce the concept.[2] Breadth of diffusion is measured using semantic distances between the abstracts of the papers introducing and adopting the concept.

Tracking the evolution of these variables over the past four decades, we document patterns that align closely with the mechanisms highlighted by the theory. The knowledge depth of newly introduced scientific concepts has increased steadily over time, while the language used to introduce them has become less accessible. In parallel, the breadth of adoption of new concepts has narrowed substantially. For example, concepts introduced in 2009 exhibit, over the first ten years after their introduction, a breadth of adoption half of a standard deviation lower than comparable concepts introduced in 1980. Placebo simulations indicate that this decline is not driven by a general increase in semantic uniformity across scientific articles over the same period.

We then use these measures to test the model's predictions directly. Ideas with higher knowledge depth are communicated using less accessible language, consistent with dilution costs rising more steeply than the benefits of bridging. Language accessibility has a large positive effect on

[2]Common words are defined as terms that appear frequently in papers published in the previous five years, reflecting language that was widely used in the scientific literature at the time.

diffusion breadth: ideas introduced using more accessible terms are adopted across a wider range of fields, even after accounting for intrinsic characteristics of the idea and its relevance to a specific field. Finally, supporting the central mechanism of the model, knowledge depth reduces diffusion, and a sizable part of this effect operates via changes in communication choices.

These patterns suggest that the realized social value of science may decline as knowledge accumulates, even when the intrinsic value of new discoveries remains unchanged. To study the long-run and policy implications of this mechanism, we embed it in a simple Romer-style growth model. As the stock of knowledge expands and new ideas build of deeper layers of prior work, scientists choose less accessible forms of communication, and diffusion across fields narrows. This interaction generates an endogenous slowdown in the growth of knowledge, even as the scientific frontier continues to advance. The model also highlights a role for institutions and policy. When researchers' incentives are aligned with the planner's objectives, communication becomes more accessible, the breadth of diffusion increases, and both the realized social value of scientific ideas and long-run growth improve.

Related Literature. This paper relates to a large literature on the social value of scientific research and its evolution over time. A central concern in this area is that aggregate research productivity appears to have declined, with rising investments in scientific research translating into progressively smaller gains in new ideas and economic growth (e.g., Jones, 2009; Bloom et al., 2020). Recent work emphasizes that the social value of science depends on both the production of new ideas and the extent to which those ideas are adopted and used. Existing studies have examined downstream use of science in applied innovation (Ahmadpoor and Jones, 2017; Yin et al., 2021), the organization of scientific research and its uptake by firms (Arora et al., 2020; Shvadron, 2023), and the role of institutions in shaping research incentives and impact (Azoulay et al., 2019; Bryan and Williams, 2021). Within this literature, our paper focuses on how *broadly* new scientific ideas diffuse. We show that the process of knowledge accumulation results in a decline in the realized social value of science as the breadth of its diffusion narrows, even when the intrinsic importance of discoveries is unchanged. Our analysis complements Ganguli et al. (2026), who show that declining similarity across inventions, which lead to weaker spillovers, accounts for a substantial share of the observed decline in research productivity. Our framework provides a micro-founded mechanism for why scientific distance becomes more constraining over time: as ideas grow more detached from basic principles, scientists optimally choose to use more specialized language, limiting the ability of researchers in other fields to understand and apply them.

A closely related strand studies communication, language, and dissemination as drivers of the diffusion of scientific knowledge. Prior work documents a long-run decline in readability and a growing reliance on technical and field-specific language (e.g., Hayes, 1992; Plavén-Sigray et al., 2017; Barnett and Doubleday, 2020), but offers limited insight into the mechanisms driving these trends or their consequences for diffusion. More recent work examines communication barriers directly. Mumtaz (2025) shows that translating scientific findings into more accessible language increases firms' use of academic research. Wang (2025) finds that commercialization incentives introduced by the Bayh-Dole Act led inventors to strategically reduce the readability of patents. Our theoretical framework provides a way to understand the long-run evolution of communication choices in science, and how these choices respond to incentives that favor recognition within fields.

Existing studies have identified several barriers that constrain the adoption of scientific ideas, including coordination failures in general-purpose technologies (Bresnahan and Trajtenberg, 1995), frictions in awareness or access to knowledge (Furman and Stern, 2011; Berkes and Nencka, 2024), and institutional and individual biases that disadvantage minoritized groups (Cheng and Weinberg, 2025). This paper emphasizes a different but complementary constraint: diffusion is limited by the extent to which ideas can be understood across domains, and by the communication choices scientists make when presenting them. The implications for welfare and policy are consistent with the view of science as a public good, emphasizing the role of institutions in shaping its social impact (Clancy et al., 2023).

Outline. The remainder of the paper is structured as follows. Section 2 presents the theoretical framework and derives its implications for communication choices and diffusion. Section 3 describes the data. Section 4 introduces the empirical measures and documents long-run trends in knowledge depth, language accessibility, and diffusion. Section 5 tests the core predictions of the model. Section 6 embeds the framework in a simple Romer-style growth model to study the long-run implications for the social value of science. Section 7 concludes.

2 Model

To guide our empirical analysis, we present a stylized model that captures key features of how new scientific ideas spread and find application across multiple domains of science.

We consider a setting in which researchers develop new scientific ideas of general-purpose nature, that is, with potential for application beyond the boundaries of their originating field.

Whether a new idea is adopted by a given domain depends on how effectively researchers in that domain can understand, process, and build upon it. We summarize these dimensions in a single metric, which we refer to as *comprehension*, capturing the extent to which adopters in a particular domain can recognize the value of an idea and understand its functioning, in a way that makes its application possible.[3] The model incorporates three intuitive properties of comprehension.

First, while new ideas have potential applications across multiple fields, the extent to which they can be comprehended is not uniform and declines as the cognitive framework of the field of origin—its objectives, background knowledge, methods, and terminology—differs from that of the potential adopter's field. We capture these differences through a measure of *scientific distance*, with more distant fields facing greater barriers to comprehension.

Second, comprehension is affected by the *knowledge depth* underlying the idea. Knowledge depth can be thought of as the size of the conceptual step required to move from basic principles to a working understanding of the idea. It depends on factors such as the amount of accumulated knowledge in the field of origin and the increasing complexity of new discoveries once simpler insights have been exhausted. A key property, for our purposes, is that knowledge depth interacts with scientific distance in shaping comprehension. Low-depth ideas—those close to self-evident statements—are broadly comprehensible, regardless of differences across domains in their cognitive frameworks. By contrast, the comprehension of high-depth ideas declines steeply with scientific distance, since understanding them requires larger conceptual steps when cognitive frameworks differ. As a result, knowledge depth amplifies the loss in comprehension that comes from scientific distance.

Third, the effect of knowledge depth on comprehension depends on how the idea is communicated. We capture communication choices through a variable, *language accessibility*, which exerts two opposing forces that vary across audiences. Greater accessibility has a *bridging effect*: it lowers barriers to comprehension, making the idea more intelligible to distant audiences. At the same time, it creates a *dilution effect*: it reduces rigor and technical depth, lowering comprehension for all audiences.[4] Knowledge depth amplifies both forces: as ideas are built on deeper layers of accumulated knowledge, both the benefits of bridging and the costs of dilution become larger.

[3]Comprehension does not require complete technical mastery. It captures a functional level of understanding sufficient to recognize an idea's potential value and apply it in a new context. The required level of understanding varies depending on the specific application. In our framework, comprehension captures this varying threshold: it reflects the probability that a researcher's understanding exceeds what is needed for adoption.

[4]Language accessibility is the only choice variable in the model. Both knowledge depth (an intrinsic property of the idea) and scientific distance (an intrinsic property of the relationship between fields) are taken as given by the researcher.

Researchers introducing new ideas are intrinsically motivated by their diffusion but have an own-field bias, valuing adoption by nearby fields more than by distant ones.[5] They select language accessibility weighing the benefits of bridging against the costs of dilution. The knowledge depth of the idea and the degree to which adoption by distant fields is discounted relative to nearby ones determines how this tradeoff is resolved.[6]

2.1 Setting

We formalize these ingredients in a static setting where a researcher develops a new scientific idea with an intrinsic degree of knowledge depth $k > 0$. There is a unit mass of potential adopters of the idea, uniformly distributed over a continuum of fields indexed by $i \in [0, 1]$. Fields are ordered so that the index i represents their scientific distance from the origin field of the idea, which we set to $i = 0$.

Each adoption generates a constant social payoff $v > 0$. The aggregate social value of the idea, V, is therefore proportional to the total measure of adopters across fields, M:[7]

$$V = v\, M = v \int_0^1 \mu(i)\, \mathrm{d}i, \tag{1}$$

where $\mu(i)$ denotes the adoption rate in field i. The researcher and adopters also derive intrinsic value from adoption, but we assume this to be negligible from a social welfare viewpoint.

2.1.1 Comprehension and adopter's decision

Each potential audience i attains a degree of comprehension of the idea $\Gamma(i) \in [0, 1]$, which captures how effectively adopters in i can understand the idea in a way that enables application. Each potential adopter is characterized by an idiosyncratic *comprehension requirement*, q, drawn from

[5] Own-field bias can arise from both institutional incentives—such as promotion systems that emphasize within-field citations and peer recognition—and cognitive factors, as researchers naturally place greater weight on feedback from audiences they interact with regularly and whose expertise they can readily evaluate.

[6] In practice, researchers may communicate the same idea through multiple outlets, such as technical papers and more accessible forms of dissemination. While this possibility can mitigate the tradeoff we model, it does not eliminate it or alter its nature. Producing accessible communication is costly in terms of time and effort, its effectiveness declines as ideas become more complex, and incentives to engage in such dissemination depend on researchers' interest in reaching audiences outside their field. While we abstract from this channel here, we discuss its role and opportunities for future research in the Conclusions. Allowing for multiple communication channels would not overturn the model's qualitative implications for the breadth of adoption and the social value of science.

[7] While the realized value of an idea may differ across fields, we abstract from this heterogeneity and assume that, ex ante, adoption in any field has the same expected social value. This corresponds to a setting in which a social planner has no information about the field-specific potential of the idea and, as a consequence, evaluates policies based on their ability to maximize overall adoption.

a uniform distribution over the interval $[0, 1]$. This threshold reflects the minimum degree of understanding the adopter needs in order to effectively use the idea for their specific objectives. Adoption occurs if $\Gamma(i) \geq q$, that is, if the adopter's comprehension of the idea exceeds their required level of understanding.[8]

Given the assumptions on the distribution of potential adopters and costs of adoption, the mass of adopters in field i is simply equal to their degree of comprehension of the idea:

$$\mu(i) = \Gamma(i). \tag{2}$$

2.1.2 Determinants of comprehension

Comprehension depends on three factors: the scientific distance (i) of the adopter's domain from the idea's origin, the idea's intrinsic knowledge depth (k), and the researcher's choice of language accessibility, which we denote by $a > 0$. As discussed above, comprehension declines with both distance and knowledge depth, and the two factors reinforce each other: depth reduces comprehension more at larger distances, and distance has a stronger effect for ideas built on deeper knowledge. Language accessibility generates two opposing forces: a bridging effect, which raises comprehension for distant fields, and a dilution effect, which reduces it for all audiences. Both forces are amplified by knowledge depth.

We capture these interactions by assuming that distance, knowledge depth, and accessibility combine into comprehension through a simple function $\gamma(i, k, a)$, defined as

$$\Gamma(i) = \gamma(i, k, a) \equiv \zeta - \lambda\, k^{\psi}\, a^{-\phi}\, i - \eta\, k\, a, \tag{3}$$

where ζ, λ, η, ψ, and ϕ are positive constants. We assume that k cannot exceed a value $k_{\max} > 0$, representing the highest hypothetical degree of knowledge depth within reach for human cognition, and impose restrictions on the other parameters so that, in equilibrium, $\Gamma(i)$ is positive for all $i \in [0, 1]$ and $k \in [0, k_{\max}]$. Equation (3) defines comprehension as a baseline ζ, capturing the intelligibility of self-evident ideas ($k = 0$), minus two additively separable terms.

The first term, $\lambda\, k^{\psi}\, a^{-\phi}\, i$, grows with both distance and knowledge depth but is mitigated by language accessibility. This term captures the loss in comprehension that arises when a "deep" idea

[8] In practice, scientific distance may also directly affect an idea's intrinsic *applicability* across domains. For example, a concept in molecular biology may be more applicable to medicine than to computer science. Our model abstracts from this dimension and focuses on comprehension (itself a function of scientific distance) as the sole driver of adoption.

is delivered to an audience with a distant cognitive framework, with the parameter λ controlling the size of this loss and the parameter ψ its elasticity with respect to depth. The *bridging effect* refers to the gain in comprehension produced via this channel by an increase in language accessibility:

$$\text{Bridging effect:} \qquad \frac{\partial(-\lambda\, k^{\psi}\, a^{-\phi}\, i)}{\partial\, a} = \phi\, \lambda\, k^{\psi}\, a^{-\phi-1}\, i > 0,$$

where the parameter ϕ controls the strength of this effect.

The second term, $\eta\, c\, a$, reflects the fact that, regardless of distance, ideas with higher knowledge depth require greater precision and technical detail to be understood in a way that enables effective use. Increasing language accessibility can reduce this precision, creating a *dilution effect* that lowers comprehension for all audiences:

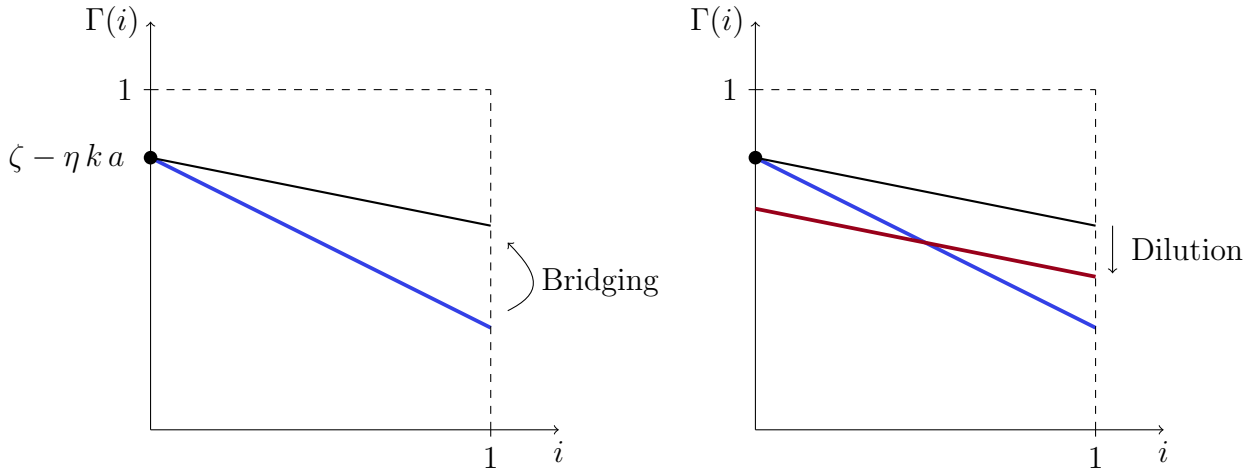
$$\text{Dilution effect:} \qquad \frac{\partial(-\eta\, k\, a)}{\partial\, a} = -\eta\, k < 0.$$

The parameter η controls the size of this loss. Note that we set log-units of k and a so that the elasticity of the second term with respect to both variables is equal to one.

Figure 1 illustrates these effects. For a given level of accessibility, comprehension declines linearly with scientific distance. Increasing a affects this relationship in two ways. The bridging effect, shown in the left panel, rotates the comprehension line around the intercept, yielding large gains in $\Gamma(i)$ for distant fields and smaller gains for nearby ones. The dilution effect, shown in the right panel, shifts the line downward, lowering comprehension for all audiences. On net, higher accessibility raises comprehension for sufficiently distant audiences while lowering it for nearby ones.

Higher knowledge depth magnifies both the costs of dilution and the benefits of bridging, as illustrated in Figure 2. The figure depicts how an increase in accessibility affects ideas of different knowledge depth. For low-depth ideas (left panel), which are close to basic principles, comprehension is broadly high and declines only modestly with distance. In this case, language plays a limited role: raising accessibility from a_{low} to a_{high} produces only small gains for distant audiences and small losses for nearby ones. By contrast, for high-depth ideas (right panel), comprehension falls steeply with distance. Here, greater accessibility generates a strong bridging effect that improves comprehension for distant fields, but also a substantial dilution effect that reduces it for nearby audiences.

Figure 1: **Effect of an increase in language accessibility: bridging and dilution**



Notes: The figure illustrates how an increase in language accessibility affects comprehension across fields at varying scientific distance, holding knowledge depth fixed. The left panel shows the bridging effect: the comprehension line rotates around the intercept. The right panel shows the dilution effect: comprehension declines for all audiences.

2.1.3 The researcher's problem

The researcher is intrinsically motivated by the adoption of the new idea but has an own-field bias, meaning that adoption by nearby fields is valued more than adoption by distant ones. Formally, the researcher's utility, U, equals the weighted mass of adopters across domains, where the weight given to each domain decays exponentially with distance:

$$U = \int_0^1 e^{-\delta i}\, \mu(i)\, \mathrm{d}i, \tag{4}$$

where $\delta \geq 0$ measures the degree of own-field bias.

After the idea is developed, the researcher chooses a level of language accessibility a to maximize utility. Given the adoption process and the utility in Equation (4), the researcher's problem can be written as

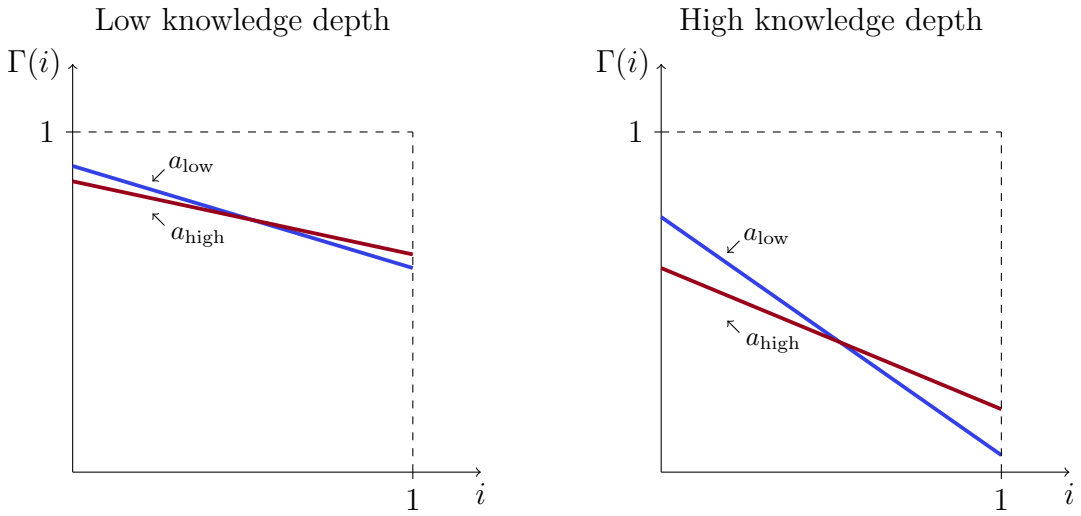
$$\max_{a>0} \int_0^1 e^{-\delta i}\, \gamma(i,k,a)\, \mathrm{d}i. \tag{5}$$

Substituting the functional form of $\gamma(i,k,a)$ into Equation (5) and taking first-order conditions yields

$$\int_0^1 e^{-\delta i}\, \phi\, \lambda\, k^{\psi}\, a^{-\phi-1}\, i\, \mathrm{d}i = \int_0^1 e^{-\delta i}\, \eta\, k\, \mathrm{d}i, \tag{6}$$

showing that the researcher sets a so that discounted bridging across all audiences (the left-hand side) equals discounted dilution (the right-hand side).

Figure 2: **How bridging and dilution vary with knowledge depth**



Notes: The figure show the effect of language accessibility on comprehension across fields at varying scientific distance, for two different levels of knowledge depth. The left panel shows this effect for a low-depth idea; The right panel for a high-depth idea. Language accessibility has a larger effect when ideas have higher knowledge depth. When concepts are closer to self-evident statements, the choice of language has a modest impact on comprehension.

2.2 Model insights

Rearranging Equation (6) gives the optimal choice of accessibility as a function of knowledge depth:

$$a^*(k) = \left(\frac{\phi \, \lambda \, \Omega(\delta)}{\eta} \, k^{\psi - 1} \right)^{\frac{1}{\phi+1}}, \tag{7}$$

where $\Omega(\delta) = \frac{1-e^{-\delta}(\delta+1)}{\delta(1-e^{-\delta})}$ is a positive term that monotonically declines with own-field bias δ, hence capturing the value the researcher derives from adoption by external audiences.

Equation (7) shows how knowledge depth affects the choice of language accessibility. The key parameter is ψ, the elasticity of the bridging effect to knowledge depth relative to the elasticity of the dilution effect (which is set to one). Intuitively, greater depth raises the benefits of accessible language, since accessibility helps convey complex ideas to audiences with distant cognitive frameworks. At the same time, depth magnifies the loss of precision and communicative efficiency that accessible language imposes on nearby audiences. When $\psi < 1$ (meaning the dilution effect is more sensitive to knowledge depth than the bridging effect) the latter force dominates, and higher knowledge depth leads the researcher to choose lower accessibility. The opposite occurs when $\psi > 1$. In summary:

Insight 1. *If the elasticity of dilution to knowledge depth is larger than that of bridging, higher*

knowledge depth leads to lower language accessibility.

The equation also shows that, for any given level of knowledge depth, stronger own-field bias (δ) lowers the researcher's choice of accessibility. This is intuitive: higher a always shifts comprehension toward distant audiences at the expense of nearby ones. When the researcher places greater weight on nearby adoption, this tradeoff makes higher accessibility less attractive.

Having characterized the researcher's choice of accessibility, we now examine its implications for total adoption,

$$M \equiv \int_0^1 \mu(i)\, \mathrm{d}i = \int_0^1 \left(\zeta - \lambda\, k^{\psi}\, a^{-\phi}\, i - \eta\, k\, a\right) \mathrm{d}i. \tag{8}$$

Plugging $a^*(k)$ into Equation (8) and evaluating the integral gives

$$M = \zeta - \xi\, \underset{+}{\Delta(\phi, \delta)}\, k^{\sigma}, \tag{9}$$

where $\xi = \lambda^{\frac{1}{\phi+1}}\, \eta^{\frac{\phi}{\phi+1}}$ and $\sigma = \frac{\phi+\psi}{\phi+1}$ are positive composite constants, and $\Delta(\phi, \delta) > 0$ is monotonically increasing in δ.[9] Equation (9) shows that greater knowledge depth reduces total adoption, and that this impact is amplified—through the choice of language accessibility—by the degree of own-field bias.

The effect of language accessibility on adoption can be better understood by decomposing M into a within-field component and a component capturing how adoption extends across fields. As we show in Appendix A.1, total adoption can be written as

$$M = \mu(0) \times \underset{+}{\Theta(D)}, \tag{10}$$

where $\mu(0)$ denotes the adoption rate in the field of origin ($i = 0$), and

$$D \equiv \frac{1}{M} \int_0^1 \mu(i)\; i\; \mathrm{d}i \tag{11}$$

is the average distance of adopters from the origin field, which we refer to as the *breadth of adoption*. The function $\Theta(\cdot)$ is increasing, as broader diffusion across fields raises total adoption for a given level of within-field uptake.

This decomposition clarifies how changes in accessibility translate into changes in adoption. For

[9] Specifically, $\Delta(\phi, \delta) = \frac{1}{2}\left(\phi\, \Omega(\delta)\right)^{-\frac{\phi}{\phi+1}} + \left(\phi\, \Omega(\delta)\right)^{\frac{1}{\phi+1}}$, where $\Omega(\delta) = \frac{1 - e^{-\delta}(\delta+1)}{\delta(1-e^{-\delta})}$. It is easy to show that $\Omega(\delta) \in (0, \frac{1}{2})$ for $\delta > 0$, is decreasing in δ, and tends to $\frac{1}{2}$ as δ approaches zero. It is also easy to show that $\Delta(\phi, \delta)$ is decreasing in $\Omega(\delta)$ for $\Omega(\delta) \leq \frac{1}{2}$. It follows that $\Delta(\phi, \delta)$ is monotonically increasing in δ for $\delta \geq 0$.

a given level of knowledge depth, higher accessibility (whose effect, in equilibrium, can be isolated by lowering own-field bias δ) reduces adoption in the origin field, $\mu(0)$, by weakening precision for nearby audiences. At the same time, Equation (9) implies that total adoption M increases as accessibility rises. Since $\mu(0)$ falls with accessibility while M rises, it must be that the increase in total adoption is driven by the term D, that is, by a *broader* diffusion of the idea across fields:

Insight 2. *Higher language accessibility increases the breadth of adoption.*

Combining Insights 1 and 2 illustrates how language accessibility mediates the effect of knowledge depth on diffusion. When the sensitivity of dilution to depth exceeds that of bridging ($\psi < 1$), greater knowledge depth leads the researcher to choose lower accessibility (Insight 1), which in turn narrows the breadth of adoption (Insight 2).

Finally, consider the overall effect of knowledge depth on the breadth of adoption once the researcher's optimal choice of accessibility is taken into account. In Appendix A.2 we show that D always decreases with complexity, and this decline is larger when the researcher's own-field bias is stronger:

Insight 3. *Higher knowledge depth lowers the breadth of adoption, and the decline is stronger when the researcher places more weight on nearby fields.*

Intuitively, as knowledge depth increases, comprehension declines more steeply with scientific distance. Because of the researcher's preferences, the equilibrium choice of accessibility is systematically tilted downward, promoting adoption in nearby domains at the expense of distant ones. When ideas are close to basic principles, this bias has little effect on diffusion, since comprehension remains high even for far audiences. As the knowledge behind new ideas becomes deeper, however, the bias becomes increasingly costly, reducing the breadth—and hence the rate—of adoption.

2.3 Implications for the social value of scientific ideas

These insights have implications for how the social returns to science evolve as knowledge accumulates.

First, as ideas grow more detached from basic principles, their social value declines. In our setting, this happens not because the intrinsic content of new ideas is less valuable, but because the depth of the knowledge on which they are built makes them less comprehensible, reducing their adoption. In this sense, the social returns to science may fall even when the intrinsic value of new ideas rises or remains unchanged.

Second, researchers' own-field bias becomes increasingly costly from a social welfare perspective as the stock of knowledge grows. Even under the socially optimal choice of accessibility—obtained by setting $\delta = 0$ in Equation (7)—breadth would fall with knowledge depth, since ideas lying on deeper knowledge are disproportionately more difficult for distant audiences to understand. Own-field bias amplifies this loss by tilting accessibility toward nearby domains, leaving adoption in distant fields inefficiently low.

We return to these implications in Section 6, where we discuss how institutional incentives for recognition within closely connected fields can shape the diffusion and social value of new ideas.

2.4 Looking ahead: From theory to data

Insights 1–3 show that knowledge depth reduces the breadth of diffusion and, under specific conditions, this effect is amplified by changes in communication choices. The key parameter is the elasticity of the bridging effect to depth (ψ) relative to the elasticity of the dilution effect (which is normalized to one). When the sensitivity of bridging is lower than that of dilution ($\psi < 1$), changes in language accessibility reinforce the effect of knowledge depth on diffusion. In the remainder of the paper, we bring these insights to the data to examine how the diffusion of new scientific ideas has evolved over recent decades and how knowledge depth and language accessibility have shaped this process.

We use data on scientific publications to identify newly introduced concepts and trace their diffusion over time and across areas of science. We show that, over the period covered by our data, the depth of knowledge on which new concepts are built has increased, while the accessibility of the language used to introduce them and their breadth of adoption have declined. We then test the key mechanisms highlighted by the model: language accessibility increases the breadth of adoption of new ideas, even after controlling for their intrinsic features, and greater depth narrows diffusion, with changes in language accounting for a substantial portion of the effect.

As a final step, we embed our setting in a simple Romer-style growth model that reproduces the empirical patterns and allows us to study the implications of knowledge accumulation and own-field bias for long-run growth and the social value of science.

3 Data

We combine data from three sources to trace the introduction and diffusion of scientific concepts across fields and over time. Our core dataset consists of publications indexed in PubMed, which we augment with field and topic classifications from OpenAlex and author disambiguation from Author-ity.

3.1 PubMed

Our core dataset is built from the PubMed Baseline Repository maintained by the National Library of Medicine (NLM).[10] The dataset includes information on more than 31 million papers published between 1781 and 2022. For each paper, we extract the date (month and year) of publication, the full abstract, language, journal, and Medical Subject Headings (MeSH) codes.[11] MeSH codes are manually assigned to each paper by librarians at the NLM with the purpose of indexing articles and facilitating searching. Each code identifies an object of study within a hierarchical classification. For our analysis, we consider the second level of the hierarchy (for example, within the highest level identifying "Anatomy," the second level includes "digestive system," "respiratory system," etc.).

3.2 OpenAlex

We collect topic and field data from OpenAlex Topics (Priem et al., 2022), which assigns to each paper categories from the All Science Journal Classification (ASJC). Building upon the classification system developed by the Center for Science and Technology Studies (CWTS), OpenAlex uses the methodology proposed by Waltman and van Eck (2012) to cluster papers into 4,521 categories based on incoming and outgoing citations. It then uses these clusters to train a machine learning procedure that uses a fine-tuned large language model (LLM) and additional paper information, such as journal names, titles, and abstracts, to map each CWTS category to an ASJC field.[12]

Contrary to MeSH codes, which identify the object of the study as it emerges from the title and abstract, the ASJC fields assigned by OpenAlex, being based on citation patterns, capture

[10] The NLM produces a baseline set of MEDLINE/PubMed records in XML format that are available for download from their ftp server. Accessed April 2022.

[11] For journals that are published quarterly, we take the first month of the quarter as the publication month.

[12] A detailed description of the procedure is provided in OpenAlex's white paper available at https://docs.google.com/document/d/1bDopkhuGieQ4F8gGNj7sEc8WSE8mvLZS/edit.

the scholarly literature in which a paper is situated. For example, a paper focused on the medical causes of HIV may be assigned overlapping MeSH codes as a paper on the social consequences of HIV, but a different ASJC field.

3.3 Author-ity

To disambiguate author names, we rely on the Author-ity dataset (Torvik and Smalheiser, 2009), which clusters author name instances into researcher identities with approximately 98% accuracy, covering all PubMed publications up to 2018. The disambiguation relies on a probabilistic model that incorporates both name-specific features (e.g., middle initials, suffixes, email addresses) and article-level metadata (e.g., coauthors, affiliations, journal, MeSH terms).

The algorithm computes pairwise similarity scores for articles sharing the same author name and applies agglomerative clustering to infer individual identities. The method relies only on internal PubMed metadata—rather than external identifiers or manual curation—making it especially suited for large-scale analysis of authorship and collaboration.

3.4 Identifying new scientific concepts

To study the characteristics of papers that introduce new ideas and how these ideas are adopted, we begin by identifying the scientific concepts that each paper brings into the literature. Following Bhattacharya and Packalen (2020), we treat as a new concept any combination of words appearing in a paper's abstract that had not appeared in the abstracts of earlier publications.

To extract such combinations from the unstructured text containing the abstract, we follow established best practices and proceed in four steps. First, we clean the text by removing numbers, words with fewer than three characters, DNA or RNA sequences, and hyperlinks. We also lemmatize[13] each word and replace instances of British English spelling with its American English equivalent. Second, we split the text into sentences and sub-sentences based on punctuation and stop words. Third, we extract all the combinations of length n of adjacent words (n-grams) obtainable from those sub-sentences, dropping possible duplicates. Fourth, we keep only the n-grams that appear in the Unified Medical Language System (UMLS) (Bodenreider, 2004), a thesaurus of medical terms curated by the National Library of Medicine, and standardize each n-gram using

[13]Lemmatization is the process of reducing words to their base or dictionary form (called the *lemma*), in a way that makes them insensitive to verb tenses, plurals, and other inflected forms of words. For example, the words *running*, *ran*, and *runs* are all lemmatized to *run*.

the corresponding entry in the UMLS. This step allows us to uniquely identify a concept, even when different terms or acronyms are used to refer to it (e.g., human immunodeficiency virus and HIV).

3.5 Embeddings

Our primary measure of diffusion breadth is based on the semantic similarity of the papers introducing and those adopting a new concept. To measure this similarity, we construct vector representations of each abstract using a large-scale embedding model.[14] Embeddings represent text as points in a high-dimensional space, where geometric proximity reflects semantic similarity. Compared to traditional approaches such as TF–IDF or bag-of-words, which rely on word frequencies and ignore context, embeddings encode the meaning of words and their relations within sentences. For example, two papers on measles vaccines will appear close to each other even if they use different terminology or stylistic conventions. This property makes embeddings particularly useful for tracking how ideas diffuse across disciplines and over time, when the language used to describe them may change.

3.6 Sample

Our final sample includes all papers in PubMed published between 1970 and 2019, with English-language abstracts and whose domain, as assigned by OpenAlex, falls within the "Health Sciences" or "Life Sciences".[15] We focus on new scientific concepts introduced between 1980 and 2009. Papers from 1970-1979 serve as a burn-in period to construct a baseline dictionary of pre-existing concepts, while papers from 2010-2019 are excluded from the identification of new concepts but included as potential adopters of earlier ones.

[14] In particular, we obtain these representations using OpenAI's `text-embedding-3-large` model, which maps each abstract to a 3,072-dimensional vector.

[15] Note that some articles are written in other languages, but still provide an abstract in English. In this case, we include them in our final sample. We only consider breakthrough papers whose document type, as it appears in PubMed, falls within one of the following categories: "Journal Article," "Research Support, N.I.H., Intramural," "Research Support, Non-U.S. Gov't," "Research Support, U.S. Gov't, Non-P.H.S.," "Research Support, N.I.H., Extramural," "Research Support, U.S. Gov't, P.H.S.," "Multicenter Study," "Comparative Study," "Twin Study." In addition to these typologies of papers, when we identify adopting papers, we also consider the following types: "Randomized Controlled Trial," "Clinical Trial," "Clinical Trial, Phase I," "Clinical Trial, Phase II," "Clinical Trial, Phase III," "Clinical Trial, Phase IV," "Evaluation Studies," "Evaluation Study," "Pragmatic Clinical Trial," "Equivalence Trial," "Adaptive Clinical Trial," "Clinical Trial, Veterinary," "Observational Study, Veterinary," "Randomized Controlled Trial, Veterinary," "Validation Study," "Comparative Study."

4 Measurement and trends

Our theory predicts that rising knowledge depth lowers the breadth of adoption, with the effect being amplified, under certain parameter conditions, by worsening language accessibility. In the remainder of the paper, we confront these predictions empirically. In this section, we introduce the empirical measures of the key objects in the model and study their evolution over time throughout the period covered by our sample. In Section 5, we will use these measures to formally test the mechanisms highlighted by our theory.

4.1 Empirical model

Throughout this section, the unit of analysis is a new scientific concept introduced between 1980 and 2009, identified as described in Section 3.4.

For each concept c, we construct two related sets of papers. First, we define the set of *breakthrough* papers, denoted by $\mathcal{B}(c)$, as those that use concept c within one year of its first appearance in the literature. These papers contributed to the initial development of the idea, either by introducing it or by adopting it very early. Second, we define the set of *adopter* papers, denoted by $\mathcal{A}(c)$, as those that use the concept within ten years of its introduction.

We estimate regressions of the form:

$$Y_c = \beta_{\mathcal{T}(c)} + X_{\mathcal{P}(c)} + W_{\mathcal{M}(c)} + Z_{|\mathcal{N}(c)|} + \epsilon_c, \tag{12}$$

where Y_c denotes a concept-level outcome—such as its knowledge depth, language accessibility, or breadth of diffusion—as defined in the remainder of this section. The coefficients $\beta_{\mathcal{T}(c)}$ capture year-specific effects based on the concept's introduction year $\mathcal{T}(c)$ and constitute our main objects of interest, as they trace how concept-level outcomes have changed over time.

Equation (12) includes a rich set of controls. The term $X_{\mathcal{P}(c)}$ is a fixed effect that flexibly accounts for heterogeneity in the outcome arising from differences in the number of breakthrough and adopter papers. Specifically, we classify each concept c into one of 1,600 categories, $\mathcal{P}(c)$, defined by the interaction of 40 bins of $|\mathcal{B}(c)|$ and 40 bins of $|\mathcal{A}(c)|$. This fixed effect absorbs variation associated with a concept's intrinsic features, such as its quality and early visibility, as well as mechanical effects driven by the size of the sets of breakthrough and adopter papers. The term $W_{\mathcal{M}(c)}$ denotes a fixed effect for the most frequent MeSH code among the breakthrough

papers associated with concept c, which we denote by $\mathcal{M}(c)$.[16] This control accounts for systematic differences across subject areas in patterns of knowledge depth, language use, and diffusion. Finally, the term $Z_{|\mathcal{N}(c)|}$ is a fixed effect for the number of breakthrough authors of concept c, whose set we denote by $\mathcal{N}(c)$, flexibly accounting for mechanical differences that may arise when aggregating author-level measures to the concept level.

The estimation sample includes all concepts introduced between 1980 and 2009 that have at least three adopter papers within ten years of their introduction. For instance, a concept first appearing in 1980 may have adopters from 1981 to 1990, while one introduced in 2009 may have adopters from 2010 to 2019. We cluster standard errors at the level of the interaction between the concept's introduction year $\mathcal{T}(c)$ and its dominant MeSH code $\mathcal{M}(c)$, allowing for arbitrary correlation of residuals within these groups.

4.2 Knowledge depth

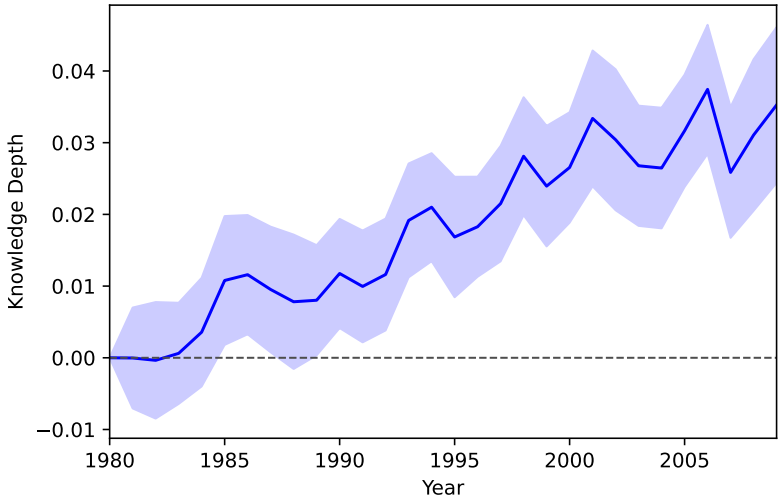
As knowledge accumulates and low-hanging fruits are exhausted, pushing the frontier requires the combination of ideas that are farther removed from basic principles. Charles Darwin could formulate the theory of natural selection largely on his own, relying on simple empirical observations and reasoning, even without any knowledge of genetics. This contrasts with the large-scale collaborations that have driven recent advances in molecular biology: the mapping of the human genome involved the coordinated work of thousands of scientists (IHGSC, 2001), and breakthroughs such as CRISPR-based gene editing and the development of AlphaFold have relied on the combined contributions of many specialized research teams.

In our framework, we take this rise in knowledge depth as the fundamental force underlying changes in diffusion. While directly measuring the knowledge depth of a concept is challenging in the absence of domain-specific expertise, we construct a concept-level measure based on a simple intuition. Following Jones (2009), who shows that deeper knowledge is reflected in higher individual specialization, we use the degree of specialization of the scientists introducing a new concept as an observable proxy. Because research built on deeper layers of prior work tends to require narrower expertise, concepts introduced by highly specialized researchers can be interpreted as stemming from deeper underlying ideas (Wuchty et al., 2007; Agrawal et al., 2016).

For each author of a breakthrough paper, we compute the Herfindahl-Hirschman Index (HHI) of the distribution of that author's publications across fields—as defined by OpenAlex—in the

[16] See the Online Appendix for details on how we assign to each concept its most frequent MeSH code.

Figure 3: **Knowledge depth (individual specialization) of newly introduced concepts: 1980-2009**



Notes: The figure plots point estimates and 95% confidence intervals of β_t from Equation (12), where the outcome K_c is defined as the average across the authors of the breakthrough papers of concept c of their individual specialization. Standard errors are clustered at the level of the concept's introduction year interacted with the primary MeSH code. $N = 39,837$. $R^2 = 0.052$.

five years preceding the breakthrough. To remove mechanical effects related to differences in publication counts, we regress the HHI on a set of fixed effects for the number of authored papers. The residual of this regression provides a measure of individual specialization. We then construct a concept-level proxy for knowledge depth, K_c, as the average individual specialization of all authors of a concept's breakthrough papers:

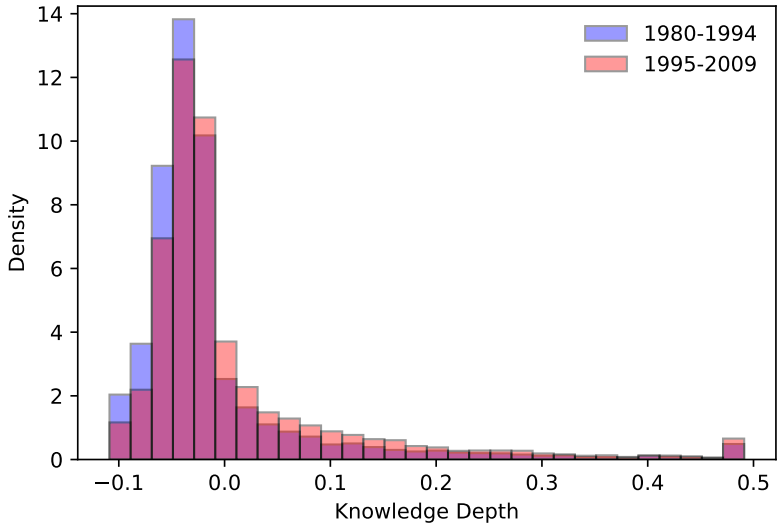
$$K_c = \frac{1}{|\mathcal{N}(c)|} \sum_{n \in \mathcal{N}(c)} \text{Specialization}_{n,\mathcal{T}(c)}.$$

Because the distribution of this measure is markedly right-skewed, we winsorize the upper tail at the 1% level.

We examine how this measure of knowledge depth has evolved over the period covered by our sample by estimating Equation (12) with concept-level knowledge depth K_c as the outcome variable. Figure 3 plots the estimated year fixed effects β_t and their 95% confidence intervals. The estimates are expressed relative to 1980, which is the omitted category (i.e., β_{1980} is normalized to zero). The plot shows a pronounced upward trend: on average, knowledge depth increases by approximately 0.04 points over the sample period, a substantial change given that the standard deviation of the measure is 0.11.[17]

[17]This finding is not itself novel, as the increase of knowledge depth over time is well documented in the literature

Figure 4: **Knowledge depth (individual specialization) of newly introduced concepts: 1980-1994 vs 1995-2009**



Notes: The figure shows the distribution of the average individual specialization of all authors of each concept's breakthrough papers. Individual specialization is computed as the Herfindahl-Hirschman Index (HHI) of the distribution of the author's publications across (OpenAlex) fields in the five years preceding the breakthrough. The measure is right-winsorized at the 1% level. Concepts are divided into an early (1980–1994, blue) and a late (1995–2009, pink) parts of the sample.

Figure 4 plots the distribution of the knowledge depth measure across all concepts, separating those introduced earlier in the sample period (1980–1994, blue) from those introduced later (1995–2009, pink). The figure shows a clear rightward shift in the distribution, with little change in its shape or dispersion. This pattern indicates that rising knowledge depth is a pervasive feature of newly introduced ideas, and is not driven by a narrow subset of high-depth concepts.

4.3 Language accessibility

Contrary to knowledge depth—which we treat as an inherent property of a new idea, determined by the state of scientific knowledge—language accessibility reflects a deliberate choice by researchers. In communicating their work, scientists balance the need for technical precision against the goal of being understood by wider audiences. How this tradeoff is resolved depends on researchers' objectives and the nature of the idea, and changes over time as the scientific landscape evolves.

Public discussions have long suggested that scientific language has become increasingly impenetrable, with frequent concerns that papers are harder to read and less accessible to non-specialists. A growing body of empirical evidence supports this view. Hayes (1992) documents a rise in lex-

on the burden of knowledge (e.g., Jones, 2009). Rather, the goal here is to confirm that our measure behaves consistently with established findings, which we take as evidence of the validity of our proxy.

ical difficulty in scientific journals throughout the twentieth century, with the largest increases occurring in general-science outlets such as *Nature* and *Science*. Plavén-Sigray et al. (2017) show that readability in scientific articles has declined steadily since the late nineteenth century, driven primarily by the growing prevalence of technical terminology. Similarly, Barnett and Doubleday (2020) document an increase in the use of acronyms in recent decades. Despite this evidence, however, the underlying drivers of this trend and its implications for how new ideas diffuse remain poorly understood.

To address this gap, we incorporate language accessibility into our empirical framework through a concept-level measure based on the prevalence of common, non-technical words in the abstracts of papers introducing new concepts. Formally, for each paper b in the set of breakthrough papers $\mathcal{B}(c)$ associated with concept c, we compute the share of words in its abstract that appear in at least 0.1% of all abstracts published between years $\mathcal{T}(c)-5$ and $\mathcal{T}(c)-1$, where $\mathcal{T}(c)$ is the year in which the concept was first introduced. The degree of language accessibility of concept c, denoted by A_c, is then defined as the median share of common words across its breakthrough papers:

$$A_c = \text{Median}\left(\left\{\frac{\#\text{ common words}_b}{\text{Total words}_b}\right\}_{b\in\mathcal{B}(c)}\right)$$

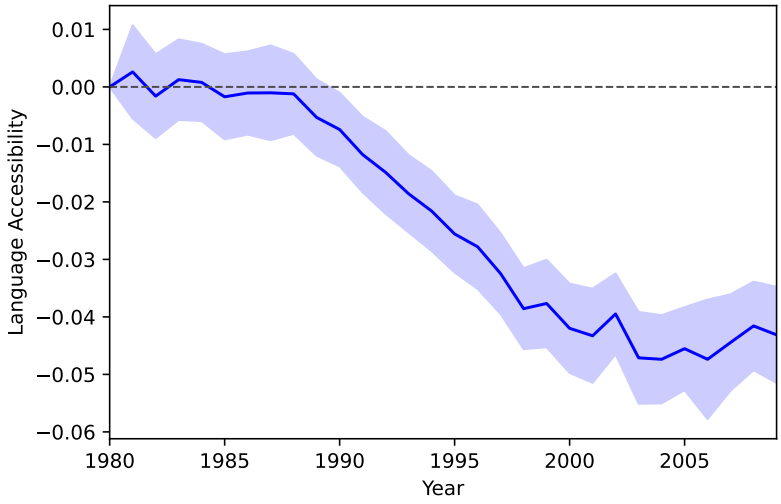
Across the concepts used in the main analysis, the average degree of accessibility is 81.4%.[18] We examine how language accessibility has evolved over time by estimating Equation (12) with concept-level accessibility, A_c, as the outcome variable.

Figure 5 reports the estimated year fixed effects β_t from 1980 to 2009, along with 95% confidence intervals. The plot reveals a steady decline in the prevalence of common words in the abstracts of papers introducing new concepts. By 2009, the share of common words had fallen by roughly 4 percentage points relative to 1980. Given that the standard deviation of the measure across the full sample is 8.8%, this represents a meaningful shift toward more specialized and less accessible scientific language.

This pattern reflects a broad change rather than the emergence of a few exceptionally impenetrable papers. Figure 6 shows the distribution of the accessibility measure across all concepts, separating those introduced early in the period (1980–1994, blue) from those introduced later (1995–2009, pink). The entire distribution shifts leftward with little change in shape or dispersion, indicating a widespread decline in language accessibility. On average, the share of common words

[18]For 40 concepts (less than 0.1% of the total), this measure equals 100%, meaning that all words in their breakthrough abstracts are classified as common according to our definition.

Figure 5: **Language accessibility of newly introduced concepts: 1980-2009**



Notes: The figure plots point estimates and 95% confidence intervals of β_t from Equation (12), where the outcome A_c is defined as the median across the breakthrough abstracts of concept c of the share of words found in at least 0.1% of the abstracts published between $\mathcal{T}(c) - 5$ and $\mathcal{T}(c) - 1$. Standard errors are clustered at the level of the concept's introduction year interacted with the primary MeSH code. $N = 42,470$. $R^2 = 0.223$.

falls from 83.4% in the early period to 79.3% in the later one.

4.4 Breadth of adoption

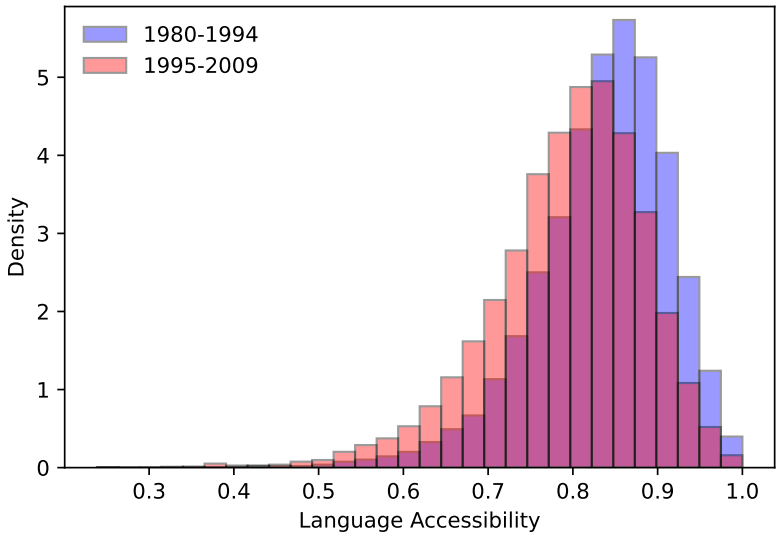
Having documented trends in the knowledge depth of new ideas and in the language used to communicate them, we now turn to their implications for diffusion. In our setting, the breadth of adoption captures how widely a new scientific concept spreads across potential users once introduced. Our theory predicts that rising knowledge depth narrows the breadth of adoption, while language accessibility can either attenuate or reinforce this effect depending on how scientists choose to communicate their work. In this subsection, we document how the breadth of adoption of new concepts has evolved over time, and in Section 5 we will study empirically how these patterns are linked to the trends in knowledge depth and language accessibility.

To quantify the breadth of adoption of a new concept, we begin by measuring the semantic distance between each pair of its breakthrough and adopter papers using the cosine distance between their abstract embeddings, which we denote by $\mathbf{e}(b)$ and $\mathbf{e}(a)$, respectively:

$$d(b, a) = 1 - \frac{\mathbf{e}(b) \cdot \mathbf{e}(a)}{\|\mathbf{e}(b)\| \|\mathbf{e}(a)\|} \qquad b \in \mathcal{B}(c),\ a \in \mathcal{A}(c).$$

For each concept c, this yields a distribution of distances across all breakthrough-adopter pairs,

Figure 6: **Language accessibility of newly introduced concepts: 1980-1994 vs 1995-2009**



Notes: The figure shows the distribution of the median across the breakthrough abstracts of each concept c of the share of words found in at least 0.1% of the abstracts published between $\mathcal{T}(c)-5$ and $\mathcal{T}(c)-1$, separating the early (1980–1994, blue) and the late (1995–2009, pink) parts of the sample.

totaling $|\mathcal{A}(c)| \times |\mathcal{B}(c)|$ combinations. We summarize this distribution with a single statistic, the concept-level breadth of adoption D_c, defined as the median of the pairwise cosine distances:

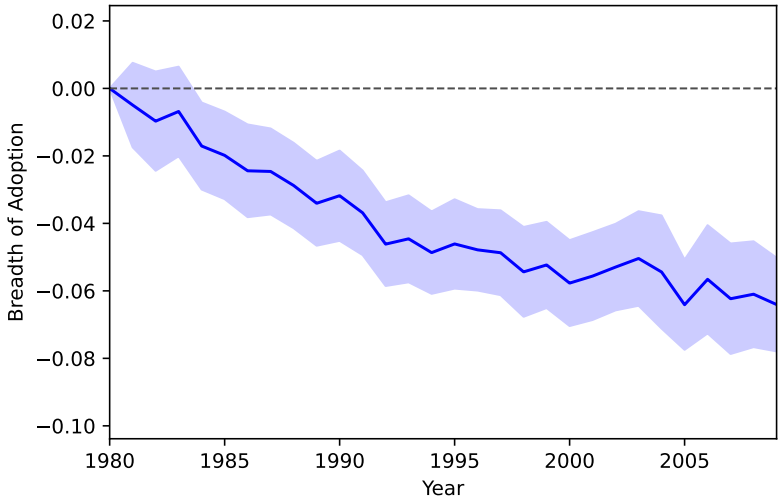
$$D_c = \text{Median}\left(\{d(b,a)\}_{b\in\mathcal{B}(c),\, a\in\mathcal{A}(c)}\right).$$

Larger values of D_c indicate that adopters are semantically farther from the original breakthroughs, corresponding to broader diffusion across scientific domains. We then estimate Equation (12) using this concept-level measure of diffusion breadth as the outcome variable.

Figure 7 plots the estimated year fixed effects β_t from 1980 to 2009, along with their 95% confidence intervals. The figure reveals a steady and pronounced decline in the semantic distance between the papers introducing a new concept and those adopting it over the subsequent decade. The magnitude of the decline is substantial: by 2009, the median cosine distance between adopters and breakthrough papers is about 0.06 lower than in 1980, corresponding to roughly half of the standard deviation of the diffusion measure in the full sample (0.126).

An important question is whether the decline in the breadth of adoption reflects a small subset of concepts with exceptionally narrow diffusion, or instead a broader shift affecting the entire distribution. Figure 8 addresses this question by plotting the distribution of diffusion measures for two equally sized time windows: 1980–1994 (blue) and 1995–2009 (pink). The figure shows a nearly uniform leftward shift in the distribution, with no visible change in its shape or dispersion,

Figure 7: **Breadth of adoption of newly introduced concepts: 1980-2009**



Notes: The figure plots point estimates and 95% confidence intervals of β_t from Equation (12), where the outcome D_c is defined as the median of the pairwise cosine distances between the breakthrough and adopters of concept k. Standard errors are clustered at the level of the concept's introduction year interacted with the primary MeSH code. $N = 42,470$. $R^2 = 0.086$.

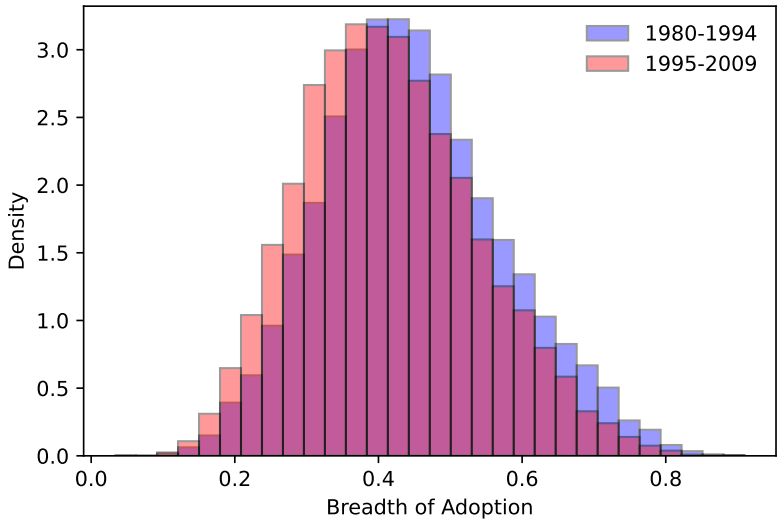
pointing to a systematic narrowing in how widely new ideas diffuse.

A potential concern with the patterns shown in Figures 7 and 8 is that the observed decline in the semantic distance between breakthrough and adopter papers may reflect a general shift in language use, objects of study, and methodologies over time rather than a change in the breadth of adoption of new ideas. Because our diffusion measure is based on cosine distances between text embeddings, increasing uniformity in those dimensions could mechanically reduce measured distances, even if the underlying spread of ideas remained unchanged.

To address this concern, we conduct a set of placebo simulations. For each concept c, we assign a set of simulated adopter papers drawn at random under the constraint that their publication year and most frequent MeSH code match those of the actual adopters. For each simulation s, we compute the corresponding diffusion measure $\hat{D}_{c,s}$ and then use these simulated measures to estimate year fixed effects β_t from Equation (12). Comparing the observed diffusion trends to these simulated benchmarks allows us to separate changes in adoption breadth from those arising from broader shifts in semantic similarity.

Figure 9 presents the results from the placebo exercise. The red line plots the average of the estimated β_t, together with their minimum and maximum, across 12 simulations. This series exhibits a mild upward trend, suggesting that scientific papers have not become more semantically uniform since 1980, consistent with analogous patterns documented for patent data by Ganguli

Figure 8: **Breadth of adoption of newly introduced concepts: 1980-1994 vs 1995-2009**



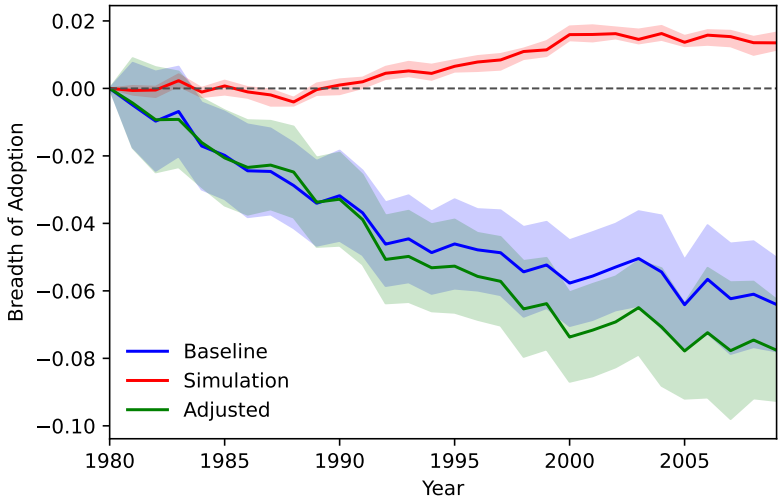
Notes: The figure shows the distribution of the median cosine distance between breakthrough and adopters across concepts, separating the early (1980–1994, blue) and the late (1995–2009, pink) parts of the sample.

et al. (2026). The green line shows the estimates of β_t when the outcome variable is defined as the adjusted diffusion measure, $\tilde{D}_c$, calculated for each concept as the difference between the actual (D_c) and average simulated ($\bar{D}_c$) breadth of adoption. This series displays a clear downward trend, closely mirroring the baseline estimates shown in blue. These results confirm that the decline in measured diffusion is not driven by general changes in language, methods, or objects of study, but instead reflects an effective reduction in the breadth of adoption of new scientific ideas.

4.5 From measurement to mechanisms

This section has documented three broad trends in the introduction and diffusion of new scientific ideas. The knowledge depth of frontier research has risen steadily over time; the language used to communicate new ideas has become less accessible; and the breadth of adoption of new concepts has declined, indicating that new ideas now spread across a narrower range of potential users. While descriptive, these patterns align closely with the mechanisms highlighted by our theory: rising knowledge depth reduces diffusion both directly and indirectly, through researchers' responses in communication choices. The next step is to examine these relationships explicitly.

Figure 9: **Breadth of adoption of newly introduced concepts: Data vs Simulations**



Notes: The blue and green lines plot point estimates and 95% confidence intervals of β_t from Equation (12). Blue line: the outcome is defined as D_c, i.e., the median of the pairwise cosine distances between the breakthrough and adopters of concept c ($R^2 = 0.086$). Green line: the outcome is defined as $\tilde{D}_c = D_c - \bar{D}_c$, that is, the difference between the median of the pairwise cosine distances between the breakthrough and adopters of concept c, and the corresponding average across 12 simulations ($R^2 = 0.112$). The red line plots the average point estimates of β_t from Equation (12) for the 12 simulations, together with the minimum and maximum estimate, where the outcome in each estimation s is defined as $\hat{D}_{c,s}$. Standard errors are clustered at the level of the concept's introduction year interacted with the primary MeSH code. $N = 42,470$.

5 Testing the model's predictions

In this section, we test the model's central predictions linking knowledge depth, language accessibility, and the breadth of adoption of new ideas. We begin by examining whether ideas lying on deeper knowledge are expressed in less accessible language, which provides evidence on whether the dilution effect is more sensitive to complexity than the bridging effect (Insight 1). Next, we test the model's prediction that language accessibility increases the breadth of adoption (Insight 2). Finally, we assess the overall effect of knowledge depth on diffusion (Insight 3) and the extent to which this effect operates through changes in language accessibility. These results connect the empirical trends documented above to the theoretical mechanisms highlighted in our model.

5.1 Ideas built on deeper knowledge are written in less accessible language

Our theory suggests that the relationship between knowledge depth and language accessibility reflects the balance of two opposing forces. Simplifying language makes ideas more comprehensible to outsiders (bridging) but reduces precision and technical rigor (dilution). Because both forces

are amplified by knowledge depth, the net effect depends on which one grows faster. If the cost of dilution rises more steeply than the benefit of bridging (i.e., if $\psi < 1$), scientists introducing ideas with greater knowledge depth will use less accessible language. Whether this condition holds in practice is an empirical question, to which we now turn.

The aggregate trends documented above, while consistent with a negative effect of knowledge depth on language accessibility, may reflect underlying factors—whether common or independent, and whether operating at the aggregate or subject-area level—influencing both variables. To isolate the direct link between them, we estimate their relationship using a specification that controls for year fixed effects interacted with the most frequent MeSH code associated with each concept:

$$A_c = X_{\mathcal{P}(c)} + \tilde{W}_{\mathcal{M}(c),\mathcal{T}(c)} + Z_{|\mathcal{N}(c)|} + \theta\, K_c + \epsilon_c. \tag{13}$$

The coefficient θ captures the effect of concept-level knowledge depth (K_c) on language accessibility (A_c), net of period-specific effects that vary across subject areas ($\tilde{W}_{\mathcal{M}(c),\mathcal{T}(c)}$), as well as flexible controls for other concept-level characteristics ($X_{\mathcal{P}(c)}$ and $\tilde{Z}_{|\mathcal{N}(c)|}$).

Table 1 reports the results, progressively adding fixed effects from Equation (13). Column 1 includes only basic concept-level controls (the number of breakthrough authors and the interaction between the number of breakthrough and adopter papers). The estimated coefficient on knowledge depth is negative and statistically significant. Quantitatively, an increase in knowledge depth of one standard deviation is associated with a decline in language accessibility of 7 percent of a standard deviation.

Adding fixed effects for the primary MeSH code in column 2 and interacting them with year fixed effects in column 3 reduces the magnitude of the coefficient by more than half. This attenuation suggests that much of the raw correlation between knowledge depth and accessibility reflects shared time trends and systematic differences across subject areas (if knowledge depth itself partly drives those differences, these controls may remove some meaningful variation). After accounting for these factors, however, the coefficient remains negative and statistically significant, indicating that, even within periods and subject areas, deeper ideas tend to be communicated in less accessible language.

Viewed through the lens of the model, these findings are consistent with the fact that the sensitivity of the dilution effect to knowledge depth exceeds that of the bridging effect (i.e., $\psi < 1$). In other words, as ideas grow more detached from basic principles, the cost of simplifying language rises more steeply than the benefit of reaching broader audiences, prompting scientists to use

Table 1: **Knowledge depth and language accessibility of newly introduced concepts**

	Language accessibility		
	(1)	(2)	(3)
Knowledge depth	-0.061***	-0.047***	-0.028***
	(0.006)	(0.005)	(0.005)
N	39,807	39,807	39,268
R^2	0.068	0.188	0.298
# brkthr authors f.e.	✓	✓	✓
# brkthr × adopter papers cat.	✓	✓	✓
Primary MeSH f.e.	×	✓	×
Primary MeSH × Year f.e.	×	×	✓

Notes: Standard errors clustered at the year-primary MeSH level in parenthesis. $^{***}p < 0.01$.

language that is more technical and less accessible.

5.2 Language accessibility increases the breadth of adoption

In our model, more accessible language facilitates the adoption of ideas by scientists who are farther from the originating field by improving comprehension and lowering barriers to use, expanding diffusion across fields. We now test this prediction empirically.

As a first step, we estimate the relationship between the adjusted breadth of adoption of a new concept, $\tilde{D}_c$, and its language accessibility, A_c:

$$\tilde{D}_c = X_{\mathcal{P}(c)} + \tilde{W}_{\mathcal{M}(c),\mathcal{T}(c)} + Z_{|\mathcal{N}(c)|} + \theta\, A_c + \epsilon_c. \tag{14}$$

A positive estimate of θ indicates that concepts expressed in more accessible language tend to diffuse more broadly. As in Equation (13), the specification includes year fixed effects interacted with the primary MeSH code associated with each concept, $\tilde{W}_{\mathcal{M}(c),\mathcal{T}(c)}$, so that comparisons are made among concepts introduced in the same year and subject area but differing in their degree of accessibility.

Table 2 reports the estimates from Equation (14) under different combinations of fixed effects. Across all specifications, the coefficient on accessibility, θ, is positive, statistically significant, and economically meaningful. In column 1, which includes only basic concept-level controls, a one-

Table 2: **Language accessibility and breadth of adoption of newly introduced concepts**

	Adjusted breadth of adoption		
	(1)	(2)	(3)
Language accessibility	0.275***	0.233***	0.171***
	(0.010)	(0.010)	(0.010)
N	39,807	39,807	39,268
R^2	0.086	0.107	0.190
# brkthr authors f.e.	✓	✓	✓
# brkthr × adopter papers cat.	✓	✓	✓
Primary MeSH f.e.	×	✓	×
Primary MeSH × Year f.e.	×	×	✓

Notes: Standard errors clustered at the year-primary MeSH level in parenthesis. $^{***}p < 0.01$.

standard-deviation increase in accessibility is associated an increase in the breadth of adoption of 19% of a standard deviation. Adding fixed effects for the primary MeSH code in column 2 and MeSH-by-year interactions in column 3 reduces the magnitude of the coefficient by roughly one third, yet it remains large and precisely estimated. As in the previous results, this attenuation indicates that part of the raw relationship reflects differences across subject areas and common time trends, but the relationship remains strong once these factors are taken into account.

While these results reveal a robust positive relationship between language accessibility and the breadth of adoption, they may still reflect unobserved characteristics of the concepts themselves. Concepts expressed in more accessible language could also be inherently more general or broadly applicable, which would independently lead to wider adoption. To address this concern, we turn to a setting that allows us to control for both the intrinsic attributes of each concept and its relevance across subject areas.

We examine how the adoption of a concept varies across fields as a function of how accessible its language is to potential adopters in those fields. For each concept c and field f (as defined by OpenAlex), we construct a measure of the accessibility of its breakthrough papers $b \in \mathcal{B}(c)$ to readers in field f as the share of words in their abstracts that appear in at least 0.1% of all abstracts

published in field f during the five years preceding the concept's introduction, $(\mathcal{T}(c)-5, , \mathcal{T}(c)-1)$:

$$A^F_{c,f} = \text{Median}\left(\left\{\frac{\#\text{ common words}_{b,f}}{\text{Total words}_b}\right\}_{b\in\mathcal{B}(c)}\right).$$

This measure captures the field-specific accessibility of concept c to field f, that is, how familiar its language is to researchers working in that field.

Using this measure, we test whether concepts diffuse more readily to fields for which the language used to introduce them is relatively more accessible. Specifically, we estimate the following specification at the concept-field level:

$$D^F_{c,f} = \alpha_c + W^F_{\mathcal{M}(c),f,\mathcal{T}(c)} + \theta\, A^F_{c,f} + \epsilon_{c,f}, \tag{15}$$

where $D^F_{c,f}$ measures the adoption of concept c in field f. The inclusion of concept fixed effects, α_c, absorbs all intrinsic characteristics of each concept—such as novelty, quality, or generality—so that identification comes from within-concept variation across fields. The specification also includes fixed effects for the interaction of the concept's primary MeSH, the receiving field, and the year of introduction, $W^F_{\mathcal{M}(c),f,\mathcal{T}(c)}$, which account for differences in the relevance of subject areas $(\mathcal{M}(c))$ to different fields (f) at the time of a concept's introduction $(\mathcal{T}(c))$.

Table 3 reports the estimates. Across all specifications, the results show that concepts diffuse more strongly to fields for which the language used to introduce them is more accessible, even after controlling for both the intrinsic characteristics of each concept and the relevance of its underlying topic to the receiving field. Column 1 uses the number of adopting papers in field f as the outcome variable. The estimates imply that increasing the share of common words from 0 to 1 raises the number of adopting papers by 2.76. This corresponds to an increase of more than half an adopting paper, on average, for a one-standard-deviation rise in field-specific accessibility (0.25). Column 2 shows that the effect is also substantial on the extensive margin: the probability that a concept is adopted by at least one paper in a field increases by more than 5 percentage points for each standard-deviation increase in accessibility. Column 3 shows that the effect remains both statistically and economically significant when the outcome is defined as the inverse hyperbolic sine of the number of adopters. Interpreting (under the usual caveats) the coefficient as a percentage-change effect, a one-standard-deviation increase in field-specific accessibility leads to an increase of more than 10 percent in the number of adopting papers.

Interpreted within our theory, these findings confirm that higher language accessibility in-

Table 3: **Field-specific language accessibility and adoption of newly introduced concepts**

	Adoption of concept k by field f		
	(1) #Adopters	(2) Any adopters	(3) IHS #Adopters
Field-specific language accessibility	2.757*** (0.077)	0.221*** (0.002)	0.437*** (0.004)
N	10,237,530	10,237,530	10,237,530
R^2	0.077	0.272	0.296
Concept f.e.	✓	✓	✓
Primary MeSH × Field × Year f.e.	✓	✓	✓

Notes: Standard errors clustered at the concept level in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

creases the breadth of adoption. By making ideas more comprehensible to researchers outside the originating field, accessible language facilitates diffusion across a wider range of audiences.

5.3 The overall effect of knowledge depth on the breadth of adoption

The results point to a configuration of parameters in which the dilution effect is more sensitive to knowledge depth than the bridging effect ($\psi < 1$). In this environment, rising knowledge depth reduces the breadth of adoption both directly and indirectly: deeper ideas are intrinsically harder for external audiences to understand and are also communicated in less accessible language, which further hinders diffusion across fields. As a result, changes in language accessibility amplify the impact of knowledge depth on diffusion.

To test this prediction, we estimate a concept-level regression in which the adjusted breadth of adoption, $\tilde{D}_c$, is the outcome and both knowledge depth, K_c, and language accessibility, A_c, are explanatory variables:

$$\tilde{D}_c = X_{\mathcal{P}(c)} + \tilde{W}_{\mathcal{M}(c),\mathcal{T}(c)} + Z_{|\mathcal{N}(c)|} + \theta_1 K_c + \theta_2 A_c + \epsilon_c. \tag{16}$$

The specification progressively adds the same set of fixed effects as in the previous concept-level regressions, including primary MeSH and MeSH-by-year interactions, ensuring that the estimates are comparable across exercises.

Table 4: **Knowledge depth, language accessibility, and breadth of adoption**

	Adjusted breadth of adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
Knowledge depth	-0.060***	-0.044***	-0.058***	-0.047***	-0.034***	-0.029***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Language accessibility		0.271***		0.230***		0.170***
		(0.010)		(0.010)		(0.010)
N	39,807	39,807	39,807	39,807	39,268	39,268
R^2	0.056	0.087	0.088	0.108	0.181	0.190
# brkthr authors f.e.	✓	✓	✓	✓	✓	✓
# brkthr × adopt. p. cat.	✓	✓	✓	✓	✓	✓
Primary MeSH f.e.	×	×	✓	✓	×	×
Primary MeSH × Year f.e.	×	×	×	×	✓	✓

Notes: Standard errors clustered at the year-primary MeSH level in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 4 presents the results. Across all specifications, knowledge depth has a negative effect on the breadth of adoption, as predicted by our theory. In column 1, which includes only basic concept-level controls, a one-standard-deviation increase in depth is associated with a decline in the breadth of adoption of approximately 5 percent of a standard deviation. Adding language accessibility in column 2 reduces the magnitude of the coefficient on complexity by close to 30 percent, indicating that language accessibility accounts for a substantial portion of the relationship between knowledge depth and diffusion.

Columns 3-6 introduce additional fixed effects for primary MeSH and MeSH-by-year interactions. As expected, these controls absorb part of the variation in both knowledge depth and accessibility, reducing the estimated coefficients. Nevertheless, the effect of depth on diffusion remains negative and statistically significant, and its magnitude continues to be meaningfully attenuated when language accessibility is included. The results therefore point to a robust negative relationship between knowledge depth and the breadth of adoption, with language accessibility acting as an important channel behind this effect.

These findings link the trends documented in Section 4 to the mechanism predicted by the model: as the frontier of knowledge expands and new ideas are built on deeper layers of prior work, scientists use less accessible language, which in turn narrows the diffusion of new ideas. The amplification pattern observed here—where declines in accessibility strengthen the impact of

knowledge depth on diffusion—is consistent with an environment in which the dilution effect is more sensitive than the bridging effect. In this configuration, rising knowledge depth reduces the social reach of new ideas both because they are intrinsically harder to understand and because they are communicated in ways that limit their adoption across fields.

6 Transmission of science in a Romer-style model

The empirical evidence presented so far supports the core mechanisms highlighted by our theory: as scientific ideas become more detached from basic principles, they are communicated in less accessible language, which limits their diffusion across fields and, in turn, reduces the social returns to new knowledge. In this section, we study these mechanisms in a simple growth framework to examine their broader implications for the evolution of scientific productivity, the value of knowledge over time, and the role of institutions and policies that govern how science is produced and disseminated.

6.1 Dynamic environment

We embed the static framework of Section 2 in a simple dynamic, Romer-style environment in which the stock of scientific knowledge, K_t, summarizes the state of the economy at time t. Time is discrete. Output, which can be interpreted as both aggregate consumption and welfare, is proportional to the stock of knowledge:

$$Y_t = K_t. \tag{17}$$

There is a continuum of scientific fields, which we assume are symmetric and populated by the same measure of scientists, whose total mass is equal to one. Differently from the static setup, we now represent fields as points uniformly distributed along the circumference of a circle. The arc distance between any two points measures their scientific distance. This representation ensures that all fields are ex-ante (and will also be ex-post) symmetric. The static model can be viewed as an "unwrapped" version of this circle, with the origin field placed at $i = 0$ and its antipode at $i = 1$. Beyond this geometric convenience, all primitives remain unchanged from the static environment.

The stock of scientific knowledge evolves according to:

$$K_{t+1} = K_t + h\, M_t\, K_t, \tag{18}$$

where $h > 0$ is an exogenous rate of arrival of new ideas and M_t denotes the total adoption of those ideas. As knowledge accumulates, the knowledge depth of new ideas increases. We represent this relationship by postulating an increasing function, $f(K)$, satisfying $f(0) = 0$ and $\lim_{K\to\infty} f(K) = k_{\max}$. A convenient functional form satisfying these properties is:

$$k_t = f(K_t) \equiv k_{\max}\left(1 - e^{-\rho K_t}\right), \tag{19}$$

where $\rho > 0$ determines the rate at which knowledge depth converges to its theoretical maximum as knowledge accumulates.

In each period, a mass h of new ideas with depth k_t is generated by scientists uniformly distributed around the circle. By the law of large numbers, each field receives the same number of ideas. Scientists face the same problem as in the static setting: they choose a level of language accessibility to maximize total discounted adoption (Equation 4), and the adoption process follows exactly the same structure as before (Equations 2 and 3). Because all fields are symmetric, the equilibrium choice of accessibility is identical across fields and follows Equation (7). Given this choice, the total adoption rate of a new idea, M_t, is the same as Equation (9):

$$M_t = \zeta - \xi\,\Delta(\phi, \delta)\,k_t^{\sigma}, \tag{20}$$

where ξ and σ are positive composite parameters, and $\Delta(\phi, \delta)$ is an increasing function of the own-field bias δ.

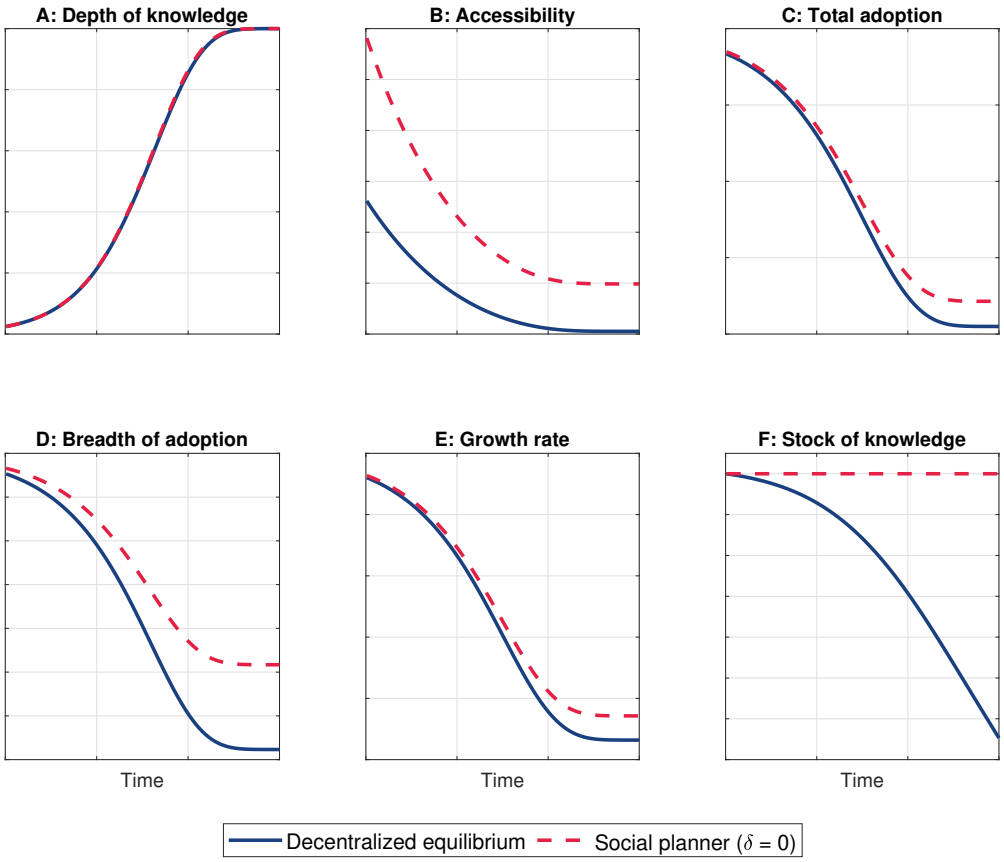
All other parameters are as defined in the static framework. The system $\{K_t, k_t, M_t\}$ fully characterizes the evolution of knowledge (and hence output) over time.

6.2 Dynamic paths and the role of policy

The model is deliberately stylized and is used to illustrate qualitative, rather than quantitative, implications of the theory. As long as the analytical restrictions derived in Section 2 hold and the empirically relevant case $\psi < 1$ is imposed, the dynamic predictions of the model do not depend on specific numerical parameter values. We therefore present the results abstracting from a particular calibration, focusing instead on trajectories that satisfy those restrictions and deliver interior solutions for language accessibility at all points in time.

Figure 10 displays the time paths of the endogenous variables. The solid blue lines correspond to the decentralized equilibrium, where scientists exhibit an own-field bias ($\delta > 0$). The dashed

Figure 10: **Dynamic paths of main variables**



Notes: The figure show the dynamics of model variables under decentralized equilibrium (solid blue, $\delta > 0$) and social planner solution (dashed red, $\delta = 0$). As knowledge depth rises, the planner chooses higher language accessibility, resulting in greater total adoption, broader diffusion across fields, and higher growth rates.

red lines correspond to the social planner solution ($\delta = 0$), in which communication choices fully internalize the social value of diffusion across fields.

As the stock of knowledge grows, the knowledge depth underlying new ideas rises monotonically (panel A). In response, scientists gradually reduce the accessibility of their language (panel B), reflecting the steepening of the tradeoff between precision for insiders and comprehension for outsiders. This decline in accessibility lowers both the overall adoption rate (panel C) and the breadth of diffusion across fields (panel D). As diffusion weakens, the growth rate of knowledge—and therefore of output—declines (panel E). The economy thus experiences an endogenous slowdown in scientific productivity: the frontier continues to expand, but each new idea contributes less to aggregate knowledge because it diffuses less broadly.

A natural point of comparison is the case in which the own-field bias is eliminated ($\delta = 0$), which, in this framework, corresponds to the planner's solution. This environment can be interpreted as one in which institutions or policies succeed in aligning scientists' incentives with the

social value of broad diffusion. When scientists place greater weight on the adoption of their ideas by other fields, they choose higher levels of accessibility, and diffusion remains substantially larger. As a result, the knowledge stock grows more quickly, and the economy converges to a balanced growth path with permanently higher growth rate and welfare, with the gap in accumulated knowledge relative to the decentralized equilibrium widening over time (panel F).

This simple setting illustrates how the interaction between rising knowledge depth and communication choices can generate a decline in the social value of new ideas and a gradual slowdown in productivity growth. It also shows how policy and institutional incentives that foster broad diffusion can mitigate these forces.

7 Discussion

We documented that scientific ideas introduced in recent decades reach audiences substantially closer to their field of origin than those from earlier periods, with the growing use of specialized language in scientific communication playing an important role in driving this trend. The mechanism we propose arises from the interaction between rising knowledge depth and scientists' language choices: as research becomes more complex, scientists optimally rely on less accessible language. Institutional incentives that lead scientists to disproportionately value recognition within their own field amplify this effect, pushing accessibility below socially optimal levels and further limiting the breadth of adoption. The empirical results support this mechanism: ideas built on deeper knowledge are communicated in less accessible language, and lower accessibility is associated with narrower diffusion.

These findings have implications for the role of science in productivity growth and for the design of research institutions and incentives. Even holding constant the number and quality of ideas produced, their contribution to productivity declines as knowledge accumulates and each idea reaches fewer potential applications. Recent work argues that as ideas become harder to find, sustaining the production of new science requires larger investments in research. Our evidence suggests suggest that narrower diffusion compounds this challenge. In this setting, policies that broaden diffusion can increase the productivity and social value of existing research without raising the rate at which new ideas are generated.

Such policies require adjustments to the incentives faced by scientists and research institutions. Existing incentive structures mainly encourage communication aimed at narrow audiences: tenure

decisions, peer review, and grant allocations are primarily designed to assess within-field impact. Several policy interventions could better align these incentives with broader diffusion. Funding agencies could place greater weight on influence beyond a paper's field of origin. Universities could incorporate measures of reach across fields into promotion decisions. Journals could adopt editorial practices that reward clarity and accessibility without sacrificing rigor. Changes in institutional incentives along these lines would encourage communication that supports wider diffusion of scientific ideas.

Our findings also highlight the value of investments in infrastructure that supports the dissemination of scientific knowledge. Science communicators who work to convey complex ideas to broader audiences are central to this process. As scientific work becomes more complex, making it accessible increasingly requires a combination of subject-matter knowledge and the ability to communicate ideas clearly to audiences of non-specialists, increasing the importance of dedicated expertise in scientific communication. New technologies, including large language models, may support research productivity by lowering the cost of translating complex ideas into forms that can be used without specialized expertise. The benefits of these efforts extend beyond adoption by other scientists. Clear and accessible communication can also strengthen public trust in science, improve the use of scientific evidence in policy design, and help inspire future generations of researchers. These broader effects point to promising avenues for future research on scientific communication and its implications for the social value of science.

References

Agrawal, A., A. Goldfarb, and F. Teodoridis (2016): "Understanding the changing structure of scientific inquiry," *American Economic Journal: Applied Economics*, 8, 100–128.

Ahmadpoor, M. and B. F. Jones (2017): "The dual frontier: Patented inventions and prior scientific advance," *Science*, 357, 583–587.

Arora, A., S. Belenzon, A. Patacconi, and J. Suh (2020): "The changing structure of American innovation: Some cautionary remarks for economic growth," *Innovation Policy and the Economy*, 20, 39–93.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2019): "Public R&D investments and private-sector patenting: evidence from NIH funding rules," *The Review of economic studies*, 86, 117–152.

Barnett, A. and Z. Doubleday (2020): "The growth of acronyms in the scientific literature," *elife*, 9, e60080.

Berkes, E. and R. Gaetani (2021): "The geography of unconventional innovation," *The Economic Journal*, 131, 1466–1514.

Berkes, E. and P. Nencka (2024): "Knowledge access: The effects of carnegie libraries on innovation," *Review of Economics and Statistics*, 1–45.

Bhattacharya, J. and M. Packalen (2020): "Stagnation and Scientific Incentives," Working Paper 26752, National Bureau of Economic Research.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020): "Are ideas getting harder to find?" *American Economic Review*, 110, 1104–1144.

Bodenreider, O. (2004): "The Unified Medical Language System (UMLS): integrating biomedical terminology," *Nucleic Acids Research*, 32(Database issue), D267–70, accessed August 2022.

Bresnahan, T. F. and M. Trajtenberg (1995): "General purpose technologies 'Engines of growth'?" *Journal of Econometrics*, 65, 83–108.

Bryan, K. A. and H. L. Williams (2021): "Innovation: market failures and public policies," in *Handbook of industrial organization*, Elsevier, vol. 5, 281–388.

Catalini, C. (2018): "Microgeography and the direction of inventive activity," *Management Science*, 64, 4348–4364.

Cheng, W. and B. Weinberg (2025): "Marginalized and overlooked? Minoritized groups and the adoption of new scientific ideas," *Journal of Labor Economics*, 43, 83–119.

Clancy, M., D. Correa, J. Dworkin, P. Niehaus, C. Watney, and H. Williams (2023): "Want to speed up scientific progress? First understand how science policy works," *Nature*, 620, 724–726.

Ekerdt, L. K. and K.-J. Wu (2024): "Self-Selection and the Diminishing Returns of Research," *Unpublished manuscript*.

Fort, T. C., N. Goldschlag, J. Liang, P. K. Schott, and N. Zolas (2025): "Growth is Getting Harder to Find, Not Ideas," *Unpublished manuscript*.

Furman, J. L. and S. Stern (2011): "Climbing atop the shoulders of giants: The impact of institutions on cumulative research," *American Economic Review*, 101, 1933–1963.

Ganguli, I., J. Lin, V. Meursault, and N. Reynolds (2026): "Mapping Inventions in Idea Space: Spreading Out Over an Expanding Frontier, 1836–2023," *Unpublished Manuscript*.

Hayes, D. P. (1992): "The growing inaccessibility of science," *Nature*, 356, 739–740.

Hill, R., Y. Yin, C. Stein, X. Wang, D. Wang, and B. F. Jones (2025): "The pivot penalty in research," *Nature*, 1–8.

IHGSC (2001): "Initial sequencing and analysis of the human genome," *Nature*, 409, 860–921.

Jones, B. F. (2009): "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?" *The Review of Economic Studies*, 76, 283–317.

Kuhn, T. S. (1962): *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1 ed.

Lane, J. N., I. Ganguli, P. Gaule, E. Guinan, and K. R. Lakhani (2021): "Engineering serendipity: When does knowledge sharing lead to knowledge production?" *Strategic Management Journal*, 42, 1215–1244.

Mumtaz, S. (2025): "Lost in Translation? Science Communication and the Commercial Diffusion of Ideas," *Unpublished manuscript*.

Myers, K. (2020): "The elasticity of science," *American Economic Journal: Applied Economics*, 12, 103–134.

Nagaraj, A., E. Shears, and M. de Vaan (2020): "Improving data access democratizes and diversifies science," *Proceedings of the National Academy of Sciences*, 117, 23490–23498.

Plavén-Sigray, P., G. J. Matheson, B. C. Schiffler, and W. H. Thompson (2017): "The readability of scientific texts is decreasing over time," *Elife*, 6, e27725.

Priem, J., H. Piwowar, and R. Orr (2022): "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," .

Shvadron, D. (2023): "Bread Upon the Waters: Corporate Science and the Benefits from Follow-On Public Research," *Unpublished manuscript*.

Torvik, V. I. and N. R. Smalheiser (2009): "Author Name Disambiguation in MEDLINE," *ACM Transactions on Knowledge Discovery from Data*, 3, 1–29.

Waltman, L. and N. J. van Eck (2012): "A new methodology for constructing a publication-level classification system of science," *Journal of the American Society for Information Science and Technology*, 63, 2378–2392.

Wang, X. (2025): "Patenting and Information Disclosure," *Available at SSRN 5387619*.

Wuchty, S., B. F. Jones, and B. Uzzi (2007): "The increasing dominance of teams in production of knowledge," *Science*, 316, 1036–1039.

Yin, Y., Y. Dong, K. Wang, D. Wang, and B. Jones (2021): "Science as a public good: Public use and funding of science," Tech. rep., National Bureau of Economic Research.

A Derivations

A.1 Derivation of Equation (10)

Consider the definition of total adoption in Equation (8):

$$M \equiv \int_0^1 \mu(i)\, \mathrm{d}i. \tag{A.1}$$

Multiplying and dividing the right-hand-side by $\mu(0)$:

$$M \equiv \mu(0) \int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i. \tag{A.2}$$

Using the fact that $\mu(i) = \zeta - \lambda\, k^{\psi}\, a^{-\phi}\, i - \eta\, k\, a$, it is easy to show that

$$\int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i = 1 - \frac{\lambda\, k^{\psi}\, a^{-\phi}}{2\,(\zeta - \eta\, k\, a)} \implies \frac{\lambda\, k^{\psi}\, a^{-\phi}}{\zeta - \eta\, k\, a} = 1 - 2 \int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i. \tag{A.3}$$

Using the same equilibrium function $\mu(i)$, we can express the breadth of adoption D as

$$D \equiv \frac{1}{M} \int_0^1 \mu(i)\; i\; \mathrm{d}i = \frac{\frac{1}{2} - \frac{\lambda\, k^{\psi}\, a^{-\phi}}{3(\zeta - \eta\, k\, a)}}{1 - \frac{\lambda\, k^{\psi}\, a^{-\phi}}{2(\zeta - \eta\, k\, a)}}.$$

Plugging Equation (A.3) into the last expression and rearranging yields

$$D = \frac{1}{4 - 6 \int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i}. \tag{A.4}$$

It is easy to show that, under our parameter restrictions, $D \in (\frac{1}{3}, \frac{1}{2})$ and, within this range, Equation (A.4) implies a monotonically incrasing relation between D and $\int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i$:

$$\int_0^1 \frac{\mu(i)}{\mu(0)}\, \mathrm{d}i = \Theta(D),$$
$$+$$

which, together with Equation (A.2), implies Equation (10).

A.2 Derivation of Insight 3

Insight 3 states that higher knowledge depth (k) lowers the breadth of adoption (D), and the decline is stronger when the researcher places more weight on nearby fields (i.e., when δ is higher or, equivalently, $\Omega(\delta)$ is lower). In other words, we will show:

$$\frac{\partial D}{\partial k} < 0 \qquad \text{and} \qquad \frac{\partial^2 D}{\partial k \, \partial \Omega(\delta)} > 0.$$

To show this, we start by using Equations (7) and (11) to express D as a function of exogenous variables and parameters only:

$$D = \frac{\frac{1}{2}\zeta - \frac{1}{3}\chi\tilde{k}\, s_1(\Omega) - \frac{1}{2}\chi\tilde{k}\, s_2(\Omega)}{\zeta - \frac{1}{2}\chi\tilde{k}\, s_1(\Omega) - \chi\tilde{k}\, s_2(\Omega)},$$

where

$$s_1(\Omega) \equiv (\phi\Omega)^{-\phi/(\phi+1)} \qquad s_2(\Omega) \equiv (\phi\Omega)^{1/(\phi+1)} \qquad \tilde{k} \equiv k^{\frac{\phi+\psi}{\phi+1}} \qquad \chi \equiv \lambda^{\frac{1}{\phi+1}}\, \eta^{\frac{\phi}{\phi+1}}.$$

Further define

$$P(\Omega) \equiv \chi\Big(\frac{1}{3}s_1(\Omega) + \frac{1}{2}s_2(\Omega)\Big), \qquad Q(\Omega) \equiv \chi\Big(\frac{1}{2}s_1(\Omega) + s_2(\Omega)\Big),$$

so that

$$D = \frac{\frac{1}{2}\zeta - P(\Omega)\tilde{k}}{\zeta - Q(\Omega)\tilde{k}}.$$

The restrictions we impose on the parameters are such that the denominator is always positive:

$$\zeta - Q(\Omega)\tilde{k} > 0.$$

Consider first the sign of $\partial D/\partial k$, which, given the definition of $\tilde{k}$, is the same as the sign of $\partial D/\partial \tilde{k}$. Differentiating D with respect to $\tilde{k}$ yields

$$\frac{\partial D}{\partial \tilde{k}} = \frac{-P(\Omega)(\zeta - Q(\Omega)\,\tilde{k}) - (\frac{1}{2}\zeta - P(\Omega)\,\tilde{k})(-Q(\Omega))}{(\zeta - Q(\Omega)\,\tilde{k})^2} = \frac{Q(\Omega)\,\frac{1}{2}\zeta - P(\Omega)\,\zeta}{(\zeta - Q(\Omega)\,\tilde{k})^2},$$

whose sign is equivalent to the sign of the numerator, $Q(\Omega)\,\frac{1}{2}\zeta - P(\Omega)\,\zeta$. It is easy to verify that

this expression is always negative:

$$\frac{1}{2}Q(\Omega) - P(\Omega) = \chi\left(\frac{1}{4} - \frac{1}{3}\right)s_1(\Omega) < 0.$$

This proves that $\frac{\partial D}{\partial \tilde{k}} < 0$ and, as a result, $\frac{\partial D}{\partial k} < 0$.

Consider now the cross-derivative $\partial^2 D/(\partial \tilde{k}\,\partial\Omega(\delta))$, which, given the definition of $\tilde{k}$, is the same as the sign of $\partial^2 D/(\partial k\,\partial\Omega(\delta))$. From the previous step we have the closed form

$$\frac{\partial D}{\partial \tilde{k}} = \frac{Q(\Omega)\,\frac{1}{2}\zeta - P(\Omega)\,\zeta}{(\zeta - Q(\Omega)\,\tilde{k})^2} = -\frac{\zeta\chi}{12}\,\frac{s_1(\Omega)}{(\zeta - Q(\Omega)\tilde{k})^2}.$$

Differentiating this expression with respect to Ω:

$$\begin{aligned}
\frac{\partial^2 D}{\partial\Omega\,\partial\tilde{k}} &= -\frac{\zeta\chi}{12}\,\frac{\partial}{\partial\Omega}\left[s_1(\Omega)\,(\zeta - Q(\Omega)\tilde{k})^{-2}\right] \\
&= -\frac{\zeta\chi}{12}\left[s_1'(\Omega)\,(\zeta - Q(\Omega)\tilde{k})^{-2} + s_1(\Omega)\cdot(-2)(\zeta - Q(\Omega)\tilde{k})^{-3}\cdot(-\tilde{k}\,Q'(\Omega))\right] \\
&= -\frac{\zeta\chi}{12}\,\frac{(\zeta - Q(\Omega)\tilde{k})\,s_1'(\Omega) + 2\tilde{k}\,s_1(\Omega)\,Q'(\Omega)}{(\zeta - Q(\Omega)\tilde{k})^3}.
\end{aligned}$$

Since $\zeta - Q(\Omega)\tilde{k} > 0$, the sign of the cross-partial is the opposite of the sign of

$$(\zeta - Q(\Omega)\tilde{k})\,s_1'(\Omega) + 2\tilde{k}\,s_1(\Omega)\,Q'(\Omega).$$

From the definition of $s_1(\Omega)$, it is trivial to show that $s_1'(\Omega) < 0$. We now show that $Q'(\Omega) < 0$ for $\Omega \in (0, 1/2)$ (the relevant range of Ω for $\delta > 0$), which, together with $s_1'(\Omega) < 0$ implies that the cross-partial derivative is positive. To show that $Q'(\Omega)$, note that

$$Q(\Omega) \propto \frac{1}{2}(\phi\Omega)^{-\frac{\phi}{\phi+1}} + (\phi\Omega)^{\frac{1}{\phi+1}},$$

so that

$$Q'(\Omega) \propto \frac{1}{2}(-\frac{\phi}{\phi+1})(\phi\Omega)^{-\frac{\phi}{\phi+1}-1}\phi + \frac{1}{\phi+1}(\phi\Omega)^{\frac{1}{\phi+1}-1}\phi.$$

Simplifying this expression, it is easy to show that $Q'(\Omega)$ is always negative for $\Omega \in (0, \frac{1}{2})$, which is the relevant range for $\delta > 0$, proving the result.