Spreading Out Across Expanding Idea Space

Ina Ganguli
UMass–Amherst

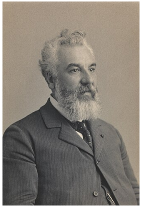
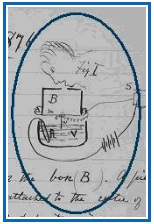
Jeffrey Lin
FRB Philadelphia

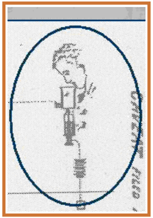
Vitaly Meursault
FRB Philadelphia

Nicholas Reynolds
University of Essex

July 2026

February 14, 1876: A Collision in Idea Space

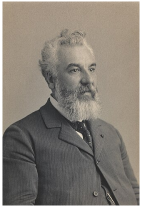


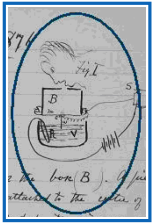

Alexander Graham Bell

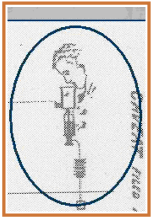
Elisha Gray

A **patent interference** — two inventors, same invention

February 14, 1876: A Collision in Idea Space





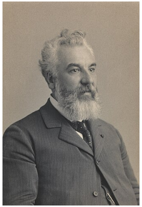


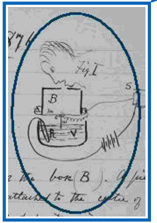
Alexander Graham Bell **Elisha Gray**

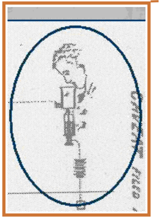
A **patent interference** — two inventors, same invention

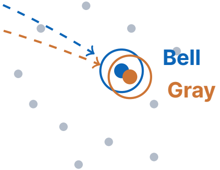
If every invention a “location” in idea space

February 14, 1876: A Collision in Idea Space









Alexander Graham Bell

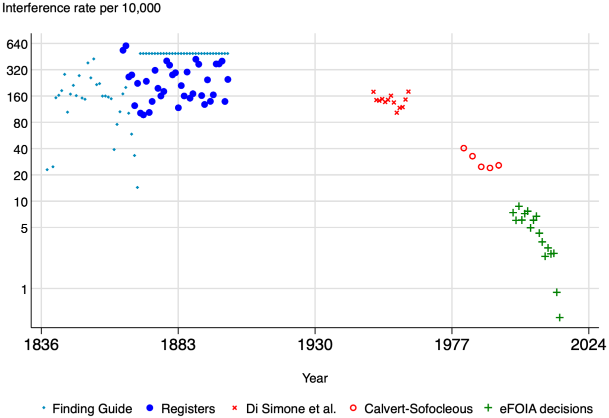
Elisha Gray

A **patent interference** — two inventors, **same invention**

They chose the **same location** in idea space

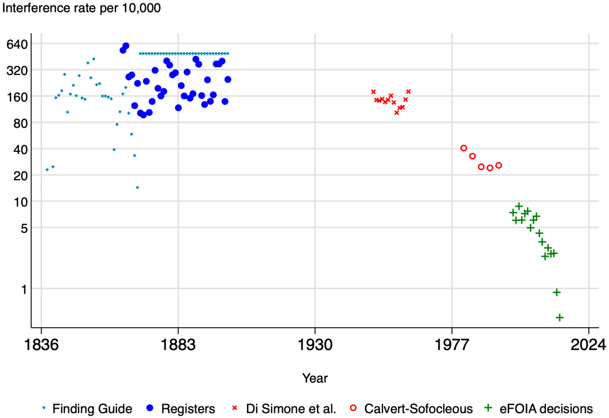
Such Collisions Were Once Routine

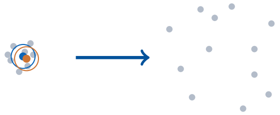
≈ 1 in 20 patents (1870s) → <1 in 10,000 (2014)!



Such Collisions Were Once Routine

≈ 1 in 20 patents (1870s) → <1 in 10,000 (2014)!





Clustered
Similar ideas

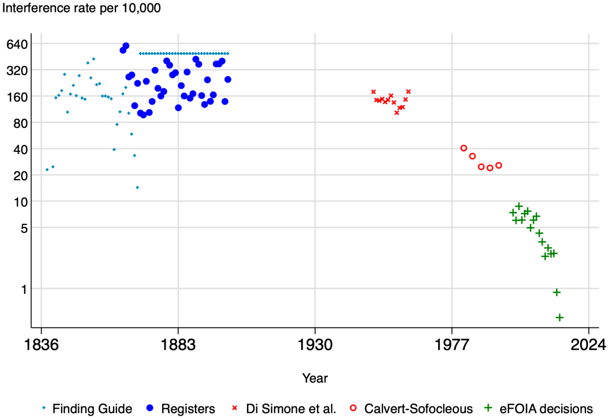
Spread Out
Dissimilar ideas

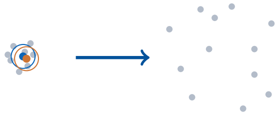
Our claim: inventors spread out across an expanding idea space

Data
Robustness

Such Collisions Were Once Routine

≈ 1 in 20 patents (1870s) → <1 in 10,000 (2014)!





Clustered
Similar ideas

Spread Out
Dissimilar ideas

Our claim: inventors spread out across an expanding idea space & **where inventors stand matters for growth**.

Data Robustness

Three Questions, Three Routes

Questions

1. Are inventions spreading out?

2. Why are they spreading out?

3. What are the consequences?

Three Questions, Three Routes

Questions

1. Are inventions spreading out?
2. Why are they spreading out?
3. What are the consequences?

How we get to the answer

1. Build & validate map of idea space

   embedding model: patent text ⟶ similarity

2. Develop theory explaining spreading

   rising entry costs + spatial competition

3. Take the model back to data

   distinctive predictions, magnitudes & decomposition

Three Questions, Three Routes

Questions

1. Are inventions spreading out?
2. Why are they spreading out?
3. What are the consequences?

How we get to the answer

1. Build & validate map of idea space

   embedding model: patent text → similarity

2. Develop theory explaining spreading

   rising entry costs + spatial competition

3. Take the model back to data

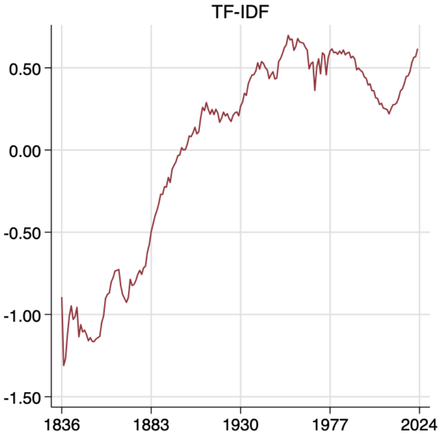
   distinctive predictions, magnitudes & decomposition

Our argument proceeds in three steps:
map the pattern → explain the mechanism → test the consequences

Are Inventions Spreading Out?

Let's Check with Text

Standardized avg pairwise cosine similarity by year



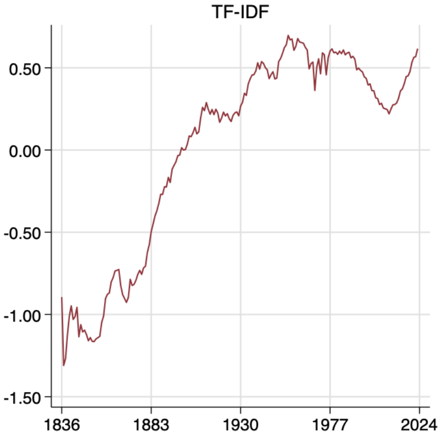
Data: Claims text of 11M+ US patents

Standard Tool: **TF-IDF** word frequencies

▸ Corpus ▸ Computation ▸ Standardization

Let’s Check with Text

Standardized avg pairwise cosine similarity by year



Data: Claims text of 11M+ US patents

Standard Tool: **TF-IDF** word frequencies

Similarity is *rising*?

× This says inventions are *crowding* — *contra* the interference record.

▸ Corpus ▸ Computation ▸ Standardization

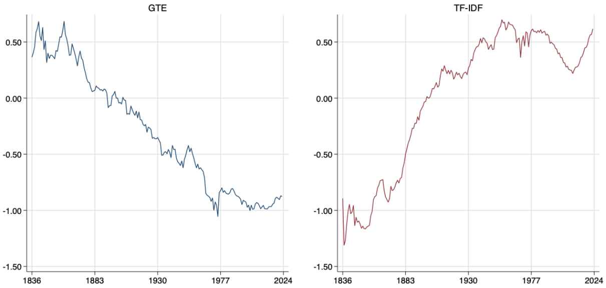
Same Text, Opposite Maps

Seven representations:

TF-IDF · GTE · PaECTER · S-BERT
Doc2Vec · USE · OpenAI

Divergent conclusions:

- TF-IDF: similarity *rising*
- Others: similarity *falling*



Which "map" should we trust?

Three Benchmarks Test Technological Similarity

No single ground truth — Validate maps against **three complementary benchmarks**:

Similarity benchmark	What it tests	Coverage
Same-invention collisions Examiner-declared interferences	**Extreme similarity:** the same patentable invention	2001–2014
Human similarity rankings Lay judgments of historical patents	**Relative proximity:** which patent pair is more similar?	1850–1975
Shared CPC categories USPTO classifications	**Categorical structure:** do two patents share a section or class?	1850–2023

A useful map should recover **legal identity**, **lay ordering**, and **administrative classifications**

Validation Selects GTE Embeddings

		Same-invention collisions	Human rankings	Shared CPC categories	
		F10	Agreement	Section ROC	Class ROC
√	**GTE**	**0.90**	**0.62**	0.596	0.656
	PaECTER	0.90	0.51	0.590	**0.672**
	S-BERT	0.82	0.54	**0.600**	0.671
×	**TF-IDF**	0.77	0.35	0.514	0.525

Higher is better in every column

√ Select GTE

Strongest historical agreement
Nearly tied on same-invention recovery
Competitive on CPC boundaries

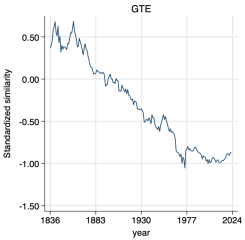
× Reject TF-IDF

Beats chance, but ranks last on every benchmark and **produces the wrong map**

▸ bicycle ≠ velocipede

Three Validated Maps Agree: Inventions Are Spreading Out

Best map: ≈1.0–1.5σ decline in similarity, 1838–2023 GTE
All 3 validated maps: ≈0.8–1.0σ decline, 1900–2000



GTE shown;
PaECTER and S-BERT exhibit
the same secular decline

Spreading out:

- ✓ after 2000 among **independent inventions**
- ✓ at **local and global spatial scales**
- ✓ **within and between** CPC classes
- ✓ **within classes as they age** and **across cohorts**

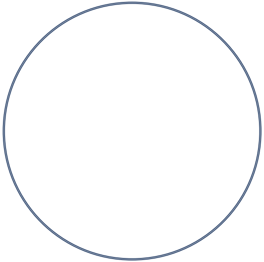
Complementary evidence:

- **Interference rates declined**
- **Embedding space expanded**

Why Are Inventions Spreading Out?

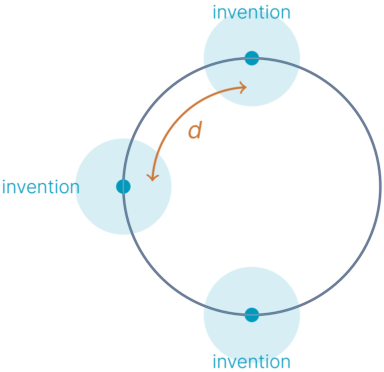
Model Setup: Spatial Competition in Idea Space

- **Idea space:** Circle of circumference H — similar problems, similar solutions



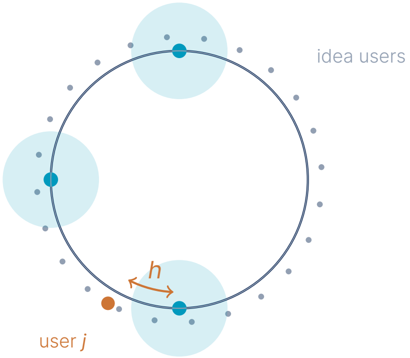
Model Setup: Spatial Competition in Idea Space

- **Idea space:** Circle of circumference H — similar problems, similar solutions
- **Idea producers:** Sell non-exclusive license to use non-rival idea. Choose **entry** f & **position**; **quality** q_i & **price** p_i. Maximize profit $\pi_i = R_i - c(q_i) - f$



Model Setup: Spatial Competition in Idea Space

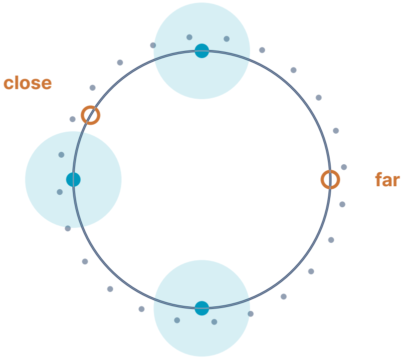
- **Idea space:** Circle of circumference H — similar problems, similar solutions
- **Idea producers:** Sell non-exclusive license to use non-rival idea. Choose **entry** f & **position**; **quality** q_i & **price** p_i. Maximize profit $\pi_i = R_i - c(q_i) - f$
- **Idea users:** Distributed uniformly on circle. Choose **one invention** to maximize **net log TFP gain** $A_j - p_i \equiv Q_i - \tau \mathbf{h} - p_i$



Model Setup: Spatial Competition in Idea Space

- **Idea space:** Circle of circumference H — similar problems, similar solutions
- **Idea producers:** Sell non-exclusive license to use non-rival idea. Choose **entry** f & **position**; **quality** q_i & **price** p_i. Maximize profit $\pi_i = R_i - c(q_i) - f$
- **Idea users:** Distributed uniformly on circle. Choose **one invention** to maximize **net log TFP gain** $A_j - p_i \equiv Q_i - \tau \mathbf{h} - p_i$

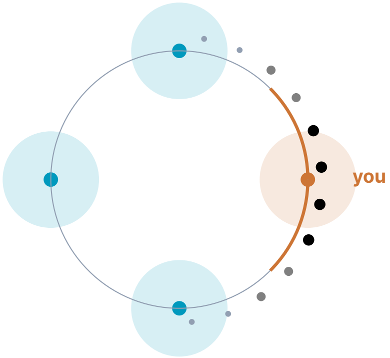
Where do you stand?



Model Setup: Spatial Competition in Idea Space

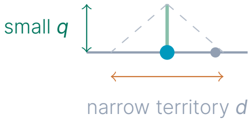
- **Idea space:** Circle of circumference H — similar problems, similar solutions
- **Idea producers:** Sell non-exclusive license to use non-rival idea. Choose **entry** f & **position**; **quality** q_i & **price** p_i. Maximize profit $\pi_i = R_i - c(q_i) - f$
- **Idea users:** Distributed uniformly on circle. Choose **one invention** to maximize **net log TFP gain** $A_j - p_i \equiv Q_i - \tau \mathbf{h} - p_i$

Where do you stand?
Isolation gives **market power**.



Wider Spacing Raises Equilibrium Quality

Close to neighboring inventor

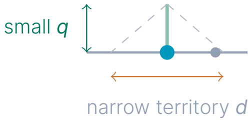


In symmetric equilibrium,

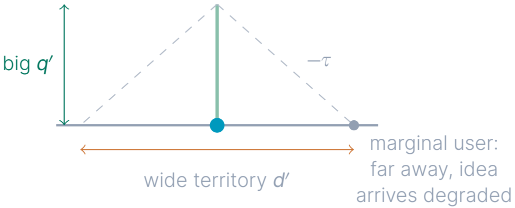
- Your idea users sit up to $d/2$ away.

Wider Spacing Raises Equilibrium Quality

Close to neighboring inventor



Far from neighboring inventor



In symmetric equilibrium,

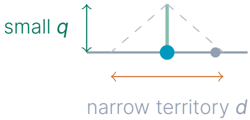
- Your idea users sit up to $d/2$ away.

The farther your marginal user:

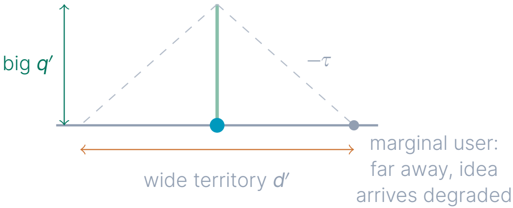
- **More pricing power**
  (users have no nearby alternatives)
- **More quality needed**
  (to be worth paying adaptation costs)

Wider Spacing Raises Equilibrium Quality

Close to neighboring inventor



Far from neighboring inventor



In symmetric equilibrium,

- Your idea users sit up to $d/2$ away.

The farther your marginal user:

- **More pricing power**
  (users have no nearby alternatives)
- **More quality needed**
  (to be worth paying adaptation costs)

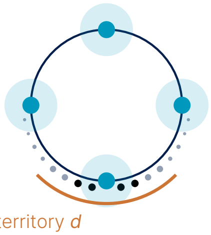
> **Spatial competition couples**
> **horizontal position**
> **& vertical investment.**

Equilibrium Turns Spacing into Prices, Quality, and Entry

1. Isolation sets your markup:

Detailed | Existence | $\tau\gamma > 1/2$ | Comparative statics | Equations

$$\boxed{p^* = \tau d}$$



Equilibrium Turns Spacing into Prices, Quality, and Entry

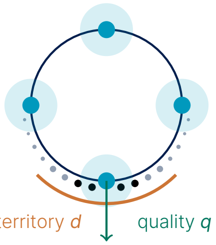
▸ Detailed ▸ Existence ▸ $\tau\gamma > 1/2$ ▸ Comparative statics ▸ Equations

1. Isolation sets your markup:

$$\boxed{p^* = \tau d}$$

2. Quality expands your territory:

$$\boxed{q^* = d/\gamma} \qquad c(q) = \tfrac{1}{2}\gamma q^{1+\eta}, \quad \eta = 1$$



Equilibrium Turns Spacing into Prices, Quality, and Entry

▸ Detailed ▸ Existence ▸ $\tau\gamma > 1/2$ ▸ Comparative statics ▸ Equations

1. Isolation sets your markup:

$$\boxed{p^* = \tau d}$$

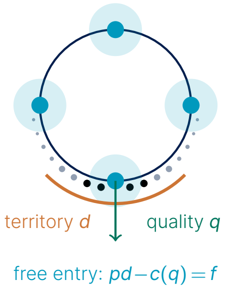
2. Quality expands your territory:

$$\boxed{q^* = d/\gamma} \qquad c(q) = \tfrac{1}{2}\gamma q^{1+\eta}, \quad \eta = 1$$

3. Territory revenue must cover R&D & entry:

$$\boxed{d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}}$$

(free entry, zero profit)



Equilibrium Turns Spacing into Prices, Quality, and Entry

▸ Detailed ▸ Existence ▸ $\tau\gamma > 1/2$ ▸ Comparative statics ▸ Equations

1. Isolation sets your markup:

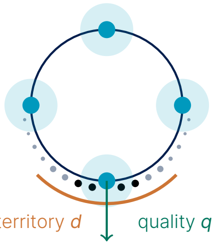
$$\boxed{p^* = \tau d}$$

2. Quality expands your territory:

$$\boxed{q^* = d/\gamma} \qquad c(q) = \tfrac{1}{2}\gamma q^{1+\eta}, \quad \eta = 1$$

3. Territory revenue must cover R&D & entry:

$$\boxed{d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}} \qquad \text{(free entry, zero profit)}$$



Geometry pins where you stand, what you charge, and how big you build.

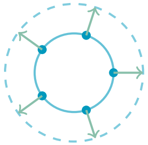
From Static Equilibrium to Self-Reinforcing Growth Loop

As innovation expands frontier… Weitzman 1998

$$\dot{H} = \delta \cdot n \cdot (Q - \tau d/4)$$

…the burden of knowledge grows… Jones 2009

$$f = \phi H^{\alpha}$$ baseline $\alpha = 1$



Only rising H, f deliver spreading out & expanding variety simultaneously

KPF | Scenarios | α

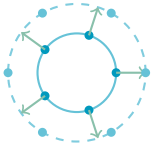
From Static Equilibrium to Self-Reinforcing Growth Loop

As innovation expands frontier… Weitzman 1998

$$\dot{H} = \delta \cdot n \cdot (Q - \tau d/4)$$

…the burden of knowledge grows… Jones 2009

$$f = \phi H^{\alpha} \qquad \text{baseline } \alpha = 1$$



…& inventors spread to recoup rising entry costs

$$d^* \text{ and } q^* = d/\gamma \uparrow \text{ as } H \uparrow \qquad \text{(from ZPC)}$$

Only rising H, f deliver spreading out & expanding variety simultaneously

▸ KPF ▸ Scenarios ▸ α

From Static Equilibrium to Self-Reinforcing Growth Loop

As innovation expands frontier… Weitzman 1998

$$\dot{H} = \delta \cdot n \cdot (Q - \tau d/4)$$

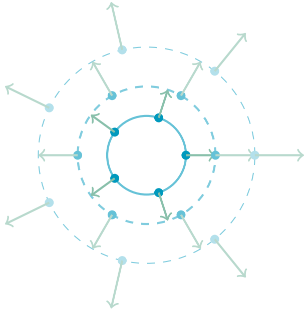
…the burden of knowledge grows… Jones 2009

$$f = \phi H^{\alpha} \qquad \text{baseline } \alpha = 1$$

…& inventors spread to recoup rising entry costs

$$d^* \text{ and } q^* = d/\gamma \uparrow \text{ as } H \uparrow \qquad \text{(from ZPC)}$$

Only rising H, f deliver spreading out & expanding variety simultaneously



The Ripple:

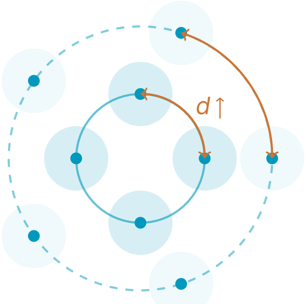
$H\uparrow \Rightarrow f\uparrow \Rightarrow d\uparrow n\uparrow q\uparrow \Rightarrow H\uparrow \Rightarrow \ldots$

▸ KPF ▸ Scenarios ▸ α

The Ripple's Fade: Attenuating Knowledge Spillovers

Gross quality includes direct diffusion from local spillovers:

$$Q_i = \underbrace{q_i}_{\text{purchased idea}} + \underbrace{\beta\left(1-\frac{d}{\lambda}\right)}_{\substack{\text{spillovers}\\ \text{(declines with } d\text{)}}} q$$



halos fade: $\beta(1-d/\lambda) \to 0$

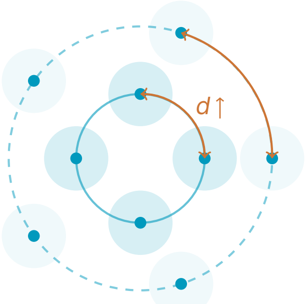
An endogenous brake on growth

Fast growth early, when inventions are packed; deceleration as they spread.

The Ripple's Fade: Attenuating Knowledge Spillovers

Gross quality includes direct diffusion from local spillovers:

$$Q_i = \underbrace{q_i}_{\text{purchased idea}} + \underbrace{\beta\left(1-\frac{d}{\lambda}\right)}_{\substack{\text{spillovers} \\ \text{(declines with } d\text{)}}} q$$



An endogenous brake on growth

Fast growth early, when inventions are packed; deceleration as they spread.

Inventors spread out, build bigger, and borrow less. cf. Fort et al. 2026

Declining R&D Productivity From Local Ideas That Are Costly to Scale

R&D productivity

$$\Pi \equiv \frac{g_{TFP}}{\text{Agg R\&D}}$$

cf. Bloom et al. 2020

Declining R&D Productivity From Local Ideas That Are Costly to Scale

R&D productivity

$$\Pi \equiv \frac{g_{TFP}}{\text{Agg R\&D}}$$

cf. Bloom et al. 2020

TFP gains are local

$$g_{TFP} \equiv \dot{\bar{A}} = \underbrace{\frac{dq}{dt}\left[1+\beta\left(1-\frac{d}{\lambda}\right)\right]}_{\text{local ideas} \times \text{spillovers}} - \underbrace{\left(\frac{\beta q}{\lambda}+\frac{\tau}{4}\right)\frac{dd}{dt}}_{\text{spreading out losses}}$$

Declining R&D Productivity From Local Ideas That Are Costly to Scale

R&D productivity

$$\Pi \equiv \frac{g_{TFP}}{\text{Agg R\&D}}$$

cf. Bloom et al. 2020

TFP gains are local

$$g_{TFP} \equiv \dot{\bar{A}} = \underbrace{\frac{dq}{dt}\left[1+\beta\left(1-\frac{d}{\lambda}\right)\right]}_{\text{local ideas} \times \text{spillovers}} - \underbrace{\left(\frac{\beta q}{\lambda}+\frac{\tau}{4}\right)\frac{dd}{dt}}_{\text{spreading out losses}}$$

R&D summed over idea space

$$\text{Agg R\&D} = \underbrace{\boxed{\mathbf{n}}}_{\text{inventions}} \cdot \underbrace{\left[\frac{1}{2}\gamma q^2 + \phi H\right]}_{\text{variable + fixed costs}}$$

▸ Decomposition

Declining R&D Productivity From Local Ideas That Are Costly to Scale

R&D productivity

$$\Pi \equiv \frac{g_{TFP}}{\text{Agg R\&D}}$$

cf. Bloom et al. 2020

TFP gains are local

$$g_{TFP} \equiv \dot{\bar{A}} = \underbrace{\frac{dq}{dt}\left[1+\beta\left(1-\frac{d}{\lambda}\right)\right]}_{\text{local ideas} \times \text{spillovers}} - \underbrace{\left(\frac{\beta q}{\lambda}+\frac{\tau}{4}\right)\frac{dd}{dt}}_{\text{spreading out losses}}$$

g_{TFP}

R&D summed over idea space

$$\text{Agg R\&D} = \underbrace{\boxed{\mathbf{n}}}_{\text{inventions}} \cdot \underbrace{\left[\frac{1}{2}\gamma q^2 + \phi H\right]}_{\text{variable + fixed costs}}$$

$n\times$ **costs**

Ideas useful only locally ⇒ TFP growth is **average** local gain
— while R&D is **sum** across an expanding idea space ⇒ Declining Π

(alongside **weakening spillovers** & traditional **fishing out and burden of knowledge** channels)

▸ Decomposition

Frontier Growth Depends on Costs, Spillovers, and Spacing

Frontier growth pinned by **cost and spillover parameters:**

$$g_H = \underbrace{\delta\left(\frac{1}{\gamma} - \frac{\tau}{4}\right)}_{\text{R\&D and adaptation costs}} + \underbrace{\frac{\delta\beta}{\gamma}\left(1 - \frac{d}{\lambda}\right)}_{\text{spillovers attenuate as inventors spread out}}$$

- Cheaper entry, R&D, adaptation, and stronger spillovers all raise growth
- **Contrast**: Prior models omit d — **where inventors stand** matters

▸ Detailed equations ▸ General model ▸ Equations

Idea-Space Geometry Endogenizes Growth Primitives

Canonical growth models sometimes treat as primitives:	Our spatial model endogenizes:
Effective spillover intensity Romer 1990	$\beta(1-d^*/\lambda)$
Innovation step size Aghion-Howitt 1992	$q^* = d^*/\gamma$
Research productivity Jones 1995	$\Pi = g_{TFP}/\text{Agg R\&D}$
Research scale Jones 1995	$n^* = H/d^*$

What Are the Consequences?

One Geometry Unifies Many Facts

Horizontal facts (d^*, n^* ↑):

√ **Spreading out**

this paper, Chiopris 2024

√ **Spillovers attenuating**

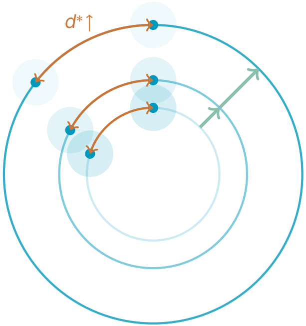
Lucking+ 19, Fort+ 26, Berkes-Gaetani 26

√ **R&D productivity declining**

Jones 95, Kortum 97, Bloom+ 20

√ **Varieties increasing**

Patent counts, Hirschey+ 12



▸ Comparative statics ▸ Kelly replication

One Geometry Unifies Many Facts

Horizontal facts (d^*, n^* ↑):

√ **Spreading out**

this paper, Chiopris 2024

√ **Spillovers attenuating**

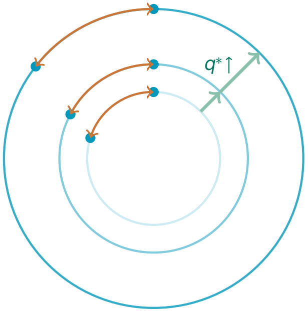
Lucking+ 19, Fort+ 26, Berkes-Gaetani 26

√ **R&D productivity declining**

Jones 95, Kortum 97, Bloom+ 20

√ **Varieties increasing**

Patent counts, Hirschey+ 12



Vertical facts (q^*, p^* ↑):

√ **R&D increasing**

Hirschey+ 12

√ **Breakthrough rates rising**

Kelly+ 2021 (horiz+vert)

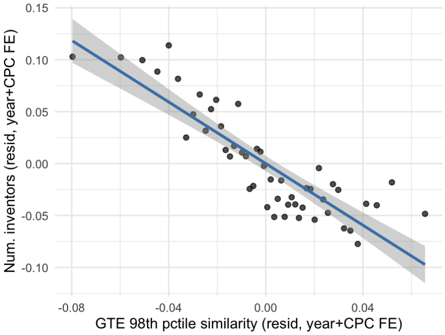
√ **Patent values rising**

Hall+ 05, Bessen+ 18, Kogan+ 17

Spreading out within fields reconciles Howitt/Peretto with Bloom+2020 evidence. Attenuating spillovers provide structural foundation for Fort+2026. Horizontal-vertical coupling reinterprets breakthrough patents (Kelly+2021).

▸ Comparative statics ▸ Kelly replication

The Model's Fingerprint: Spacing Couples with Quality



Co-inventors vs. similarity

residualized on year × CPC class

This model's distinctive prediction:

$$\boxed{q^* = d/\gamma}$$

More isolated inventions:

- Use **larger teams**
- More often **firm-assigned**
- More often **cited**
- Have **higher market value**

(KPSS: ≈\$22–33M per σ of isolation)

Holds in cross-section (220K patents, within class–year) and across the **188-year** class panel (CUSP).

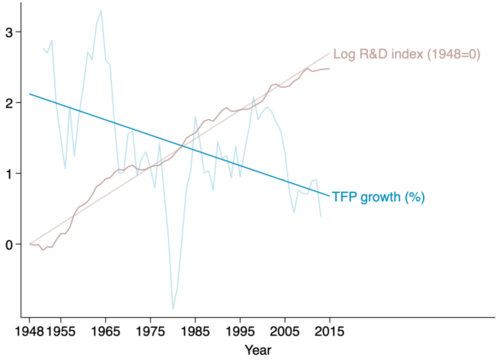
▸ All four proxies ▸ Regression table

Research Productivity Falls Through Two Margins

$$g_{\Pi} = \underbrace{g_{g_{TFP}}}_{-1.6\%/\mathrm{yr}} - \underbrace{g_{R\&D}}_{+4.0\%/\mathrm{yr}} = \boxed{-5.6\%/\mathrm{yr}}$$

How much of the **TFP-growth slowdown** comes from spreading?

How much of **R&D growth** comes from spatial forces?

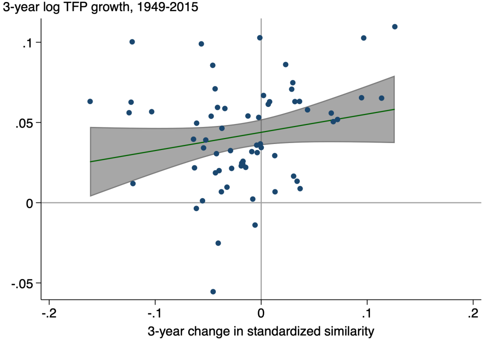


Bloom, Jones, Van Reenen and Webb, 2020

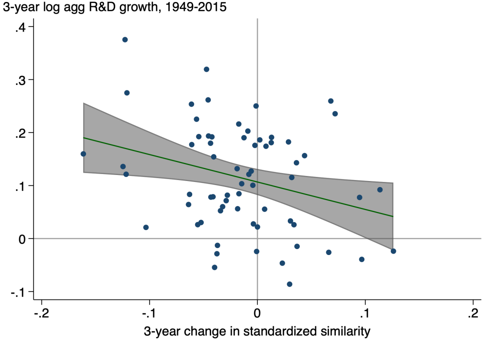
Strategy: Quantify the spatial contribution to each margin, then add them.

The Data Supply Two Moments

TFP margin



R&D margin



More spreading $\Longleftrightarrow$ slower TFP growth

More spreading $\Longleftrightarrow$ faster aggregate R&D

Spreading growth: $g_d \approx -\Delta\overline{\mathrm{Sim}}_t$ **Next:** the model interprets these comovements

How the Two Moments Identify the Decomposition

TFP moment

$$\Delta \log(\text{TFP})_t = b_0 + b_1(-\Delta\overline{\text{Sim}}_t) + b_2(-\Delta\overline{\text{Sim}}_t)(-\overline{\text{Sim}}_{t-1}) + b_3 t + \epsilon_t$$

More spreading ⟷ slower TFP growth:
Slope evaluated at observed spacing
⇓

Spillover drag = -0.16 pp/yr

1981–2001 spreading, at 1991 spacing
1948–2015 spreading: -0.11 pp/yr
directly identified from the TFP regression

R&D moment

▸ TFP bridge ▸ R&D bridge

How the Two Moments Identify the Decomposition

TFP moment

$$\Delta \log(\mathrm{TFP})_t = b_0 + b_1(-\Delta\overline{\mathrm{Sim}}_t) + b_2(-\Delta\overline{\mathrm{Sim}}_t)(-\overline{\mathrm{Sim}}_{t-1}) + b_3 t + \epsilon_t$$

More spreading $\longleftrightarrow$ slower TFP growth:
Slope evaluated at observed spacing
$\Downarrow$

Spillover drag = -0.16 pp/yr

1981–2001 spreading, at 1991 spacing
1948–2015 spreading: -0.11 pp/yr
directly identified from the TFP regression

R&D moment

$$\Delta \log(\mathrm{R\&D})_t = a_0 + a_1(-\Delta\overline{\mathrm{Sim}}_t) + \epsilon_t$$

Spacing–R&D comovement:
primarily rising quality per invention

$\Rightarrow$ slope a_1 $\Rightarrow$ variable cost share $\hat{\theta} = 72\%$

R&D growth net of spreading:
primarily more inventions & rising entry costs

$\Rightarrow$ intercept a_0 $\Rightarrow$ frontier growth $\hat{g}_H = 2.67\%$/yr

▸ TFP bridge ▸ R&D bridge

Four Quantitative Checks Support the Decomposition

Check	Our Estimate	Benchmark
External validation		
Spillover drag validates TFP regression	-0.16 **pp/yr** TFP regression	-0.15 **pp/yr** ✓ Bloom+2013 spillover elasticity × our spreading: $0.206\text{pp} \times 1.04/\sigma \times -0.007\sigma/\text{yr}$
Variable-cost share validates R&D regression	$\widehat{\theta} = 72\%$ R&D regression	69% ✓ NSF BERD R&D labor share
Internal consistency		
Idea-space growth overidentified: two estimates agree	$\widehat{g}_H = 2.67\%/\text{yr}$	$g_H = 2.7\%/\text{yr}$ ✓ implied by $g_{R\&D} = \frac{3}{2} g_H$
Spacing vs. R&D growth model ratio holds in data	$g_d / g_{R\&D} = 1/3$ model	0.2–0.4 ✓ data

▸ Spillover calculation ▸ TFP table ▸ R&D table

Spatial Forces Account for 40–60% of Declining R&D Productivity

TFP margin

Long-run spatial drag
0.11 **pp/yr**

1948–2015

full period estimate

R&D decomposition: four calibrated components

Baseline: $\theta = 0.69$, $g_H = 2.7\%/\text{yr}$, $\alpha = 1$, $\eta = 1$

Component	Calibrated Expression	pp/year
Entry expansion	$(1 - \frac{\alpha}{2})g_H$	1.33
Quality scaling	$(\theta\,\frac{\alpha}{2})g_H$	0.92
Spatial subtotal		**2.25**
Fishing out	$(\theta\eta\,\frac{\alpha}{2})g_H$	0.92
Burden of knowledge	$(1-\theta)\alpha g_H$	0.83

$$\frac{\overbrace{0.11}^{\text{TFP}} + \overbrace{2.25}^{\text{R\&D}}}{5.6} = \boxed{42\%}$$

Disciplined sensitivity (α, η)

$$\boxed{42\text{–}58\%}$$

▸ Detailed accounting ▸ Sensitivity

Three Takeaways

1 **Inventions spread out**

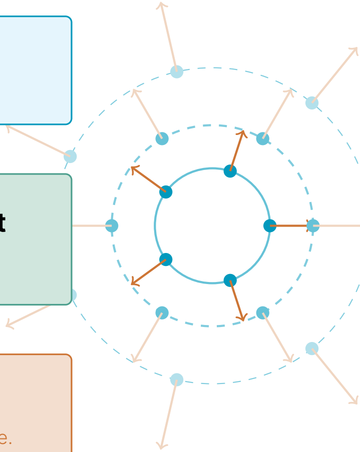
Validated NLP measures spanning 1836–2023; Interferences 1870–2014.

2 **An expanding idea frontier drives inventions apart**

Unifies empirical patterns; Endogenizes core growth primitives.

3 **Where inventors stand matters for growth**

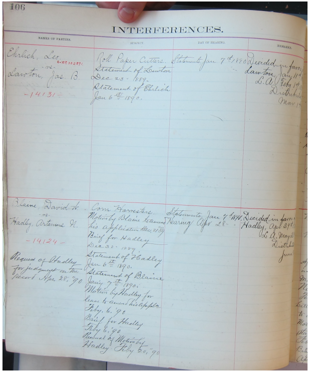
Spatial forces account for 40–60% of the long-run R&D productivity decline.



Backup Slides

Backup: Interference Data

◂ Return



Example (Jan 7, 1890)
Ehrlich v. Lawton: Roll paper cutters

Blaine v. Hadley: Corn harvesters

Purpose-digitized from National Archives:

- USPTO *Registers of Interferences*, 1864–1900
- 19,388 interference cases documented
- Average 504/year (2.7–4.9%)

Other sources:

- Case Files (1838–1900): 87–880/year
- DiSimone et al. (1950–1962): 640/year (1.4%)
- Calvert & Sofocleous (1980–1994): 237/yr (0.3%)
- Ganguli et al. (1998-2014): 76/year (0.05%)

Backup: Patent Corpus Details

◂ Return

Historical (1836–1975): ProQuest Patents Core

- OCR-digitized patent images
- Full text of claims extracted
- Quality varies with original document condition

Modern (1976–2023): USPTO PatentsView

- Machine-readable full text
- Structured data with claim parsing
- Consistent quality

Potential discontinuity at 1976:

- Some evidence of break in levels
- Trends consistent across periods
- Results robust to excluding transition years

Backup: Computing Similarity Efficiently

◂ Return

Challenge: $O(N^2)$ pairwise comparisons infeasible for millions of patents

Solution: For unit-normalized vectors, average cosine similarity reduces to:

$$\bar{S} = \frac{1}{N(N-1)} \sum_{i \neq j} \cos(v_i, v_j) = \frac{\|\sum_i v_i\|^2 - N}{N(N-1)}$$

Complexity: $O(N \cdot d)$ where d = embedding dimension

Implementation:

1. Normalize all vectors to unit length
2. Sum vectors: $S = \sum_i v_i$
3. Compute $\|S\|^2$
4. Apply formula

Cross-sectional SD: Subsample up to 10,000 patents/year

Backup: Alternative Normalizations

Return

Main specification: Standardize by annual cross-sectional SD

Robustness checks:

1. Time-invariant global SD → Nearly identical results
2. Raw similarity (no standardization) → Same qualitative pattern
3. Different sample sizes per year → Robust

Why standardize?

- Different representations have different scales
- No intrinsic economic interpretation of raw similarity
- SD provides meaningful units for comparison

Backup: Interference Validation Task

Return

Patent interferences (2001–2014):

- **First to invent**: USPTO proceeding for multiple applicants w/ identical claims
- Provides ground truth for "identical" similarity
- 322 true interfering pairs among 96,580 application pairs

Economic intuition: Examiner ranks pairs by similarity, investigates above threshold

- Higher threshold → fewer false positives but miss true interferences
- Lower threshold → catch more but burden staff with unnecessary investigations

Metrics:

- F10: Weights recall 10× more than precision (missing interferences is costly)
- PR AUC: Precision-Recall area under curve across all thresholds

Key result: GTE, PaECTER, OpenAI retrieve ~90% of true interferences with 2–5× fewer false positives than TF-IDF/S-BERT

Backup: Human Annotation Task

Return

Historical patents (1850–1975):

- Sampled patent pairs that each model ranked at least 50 percentiles apart
- Annotators rank **relative** similarity of 2 patent pairs
- Tests temporal robustness (historical language): Oversample 1880–1920

Task: Do model rankings agree with human rankings?

- For each patent, rank others by similarity
- Compare model ranking to human ranking

Metric: Agreement coefficient from regression

$$\text{Human Rank} = \alpha + \beta \cdot \text{Model Rank} + \epsilon$$

Higher β = better agreement

Backup: Classification Validation Task

Return

USPTO Classifications (1850–2023):

- CPC technology codes assigned by examiners
- Section level (8 categories) and Class level (120+ categories)
- Captures expert judgment of technological relatedness

Task: Predict whether patent pair shares classification

- Same Section (coarse): 8 top-level categories
- Same Class (fine): 3-digit classification

Metric: ROC AUC

- Area under Receiver Operating Characteristic curve
- 0.5 = random, 1.0 = perfect

Backup: Why TF-IDF Fails

Return

TF-IDF overweights period-specific language:

- Treats “velocipede” (1880s) and “bicycle” (modern) as unrelated
- Period-specific terminology dominates similarity scores
- Creates spurious correlation with time

Example: 1880 velocipede patent

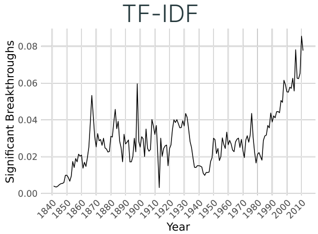
- TF-IDF: High similarity to other 1880s patents (shared vocabulary)
- GTE: High similarity to modern bicycle patents (shared concepts)

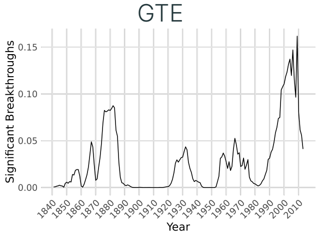
Evidence:

- TF-IDF similarity correlates with word overlap
- GTE similarity correlates with conceptual similarity
- Google Ngrams shows vocabulary shifts over time

Backup: Kelly et al. Breakthrough Replication

Return





Kelly et al. (2021): Identify "breakthrough" patents using similarity to future patents

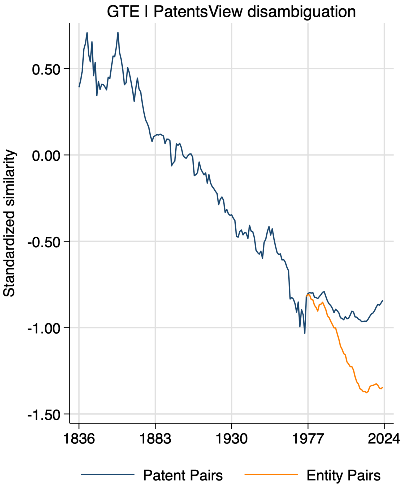
Our replication with GTE:

- Qualitative conclusions align (more breakthroughs today)
- Quantitative results more robust (less sensitivity to methodological choices)
- TF-IDF produces noisier breakthrough classification

Implication: Validated similarity measures improve downstream analyses

Backup: Robustness — Multi-Patent Entities

◂ Return



Concern: Post-2000 dynamics coincide with: business method patents, non-practicing entities, increased defensive patenting. ▸ Patents v entities

- **Multiple patents from same entity may be similar but not independent.**

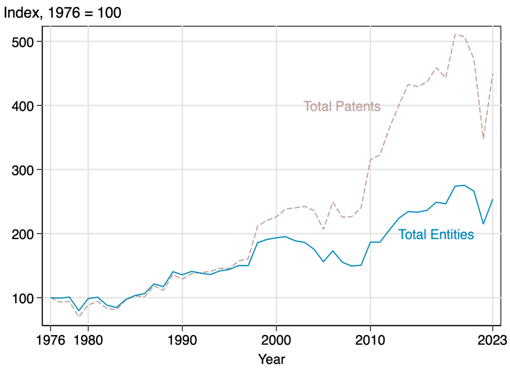
Strategy: Sample 1 patent/entity–year

Result:

- Decline persists after correction
- **Independent inventions** still spreading out

Backup: Growth in Patents vs. Patenting Entities

◂ Return



- Number of issued utility patents and unique patenting entities per year
- Divergence after 1999: substantial growth in patents per entity
- Driven by business method patents and non-practicing entities
- Motivates sampling 1 patent per entity per year for robustness

Backup: Robustness — Spreading Out Within Technology Classes

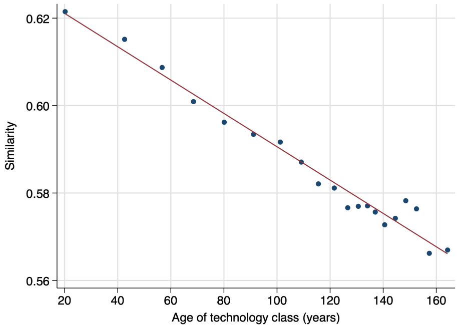
◂ Robustness

Alternative explanations: Changing patent office practice over time? Shifts across major technology areas?

Test: Within-class similarity by class “age”

- Birth = Class first issued 50 patents
- e.g., Combinatorial Chemistry 2001
- Addresses compositional concerns ▸ Between

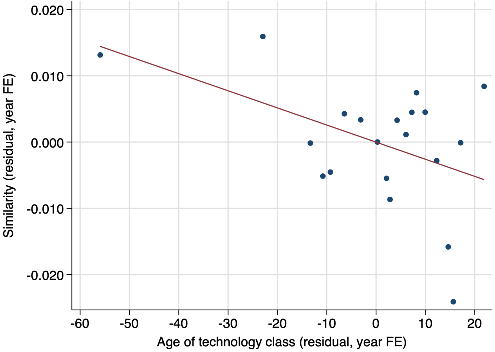
Finding: Within-class similarity declines as classes mature ▸ Class+Year FE

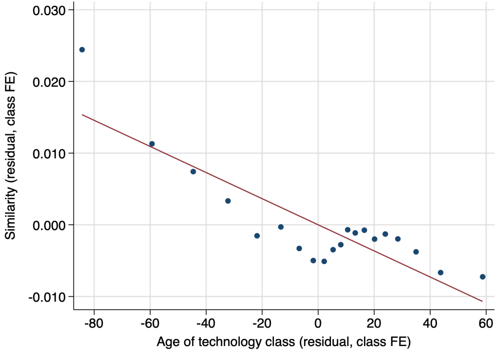


Spreading out is a dynamic process tied to field evolution

Backup: Similarity Within Class by Class Age

Return



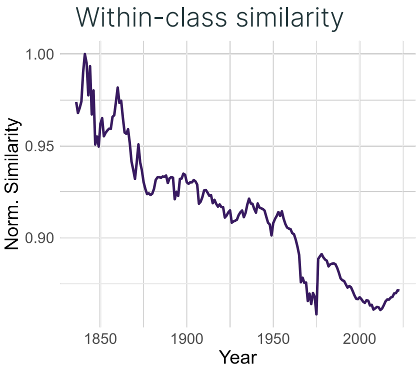


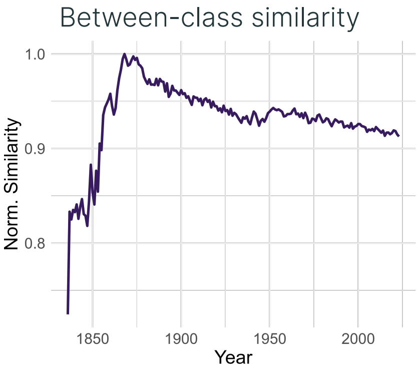
Right: Cross-cohort comparison (year FE) — robustness to patent office practice

Left: Within-cohort comparison (class FE) — robustness to composition

Backup: Within vs. Between Technology Classes

◂ Within classes ◂ Robustness

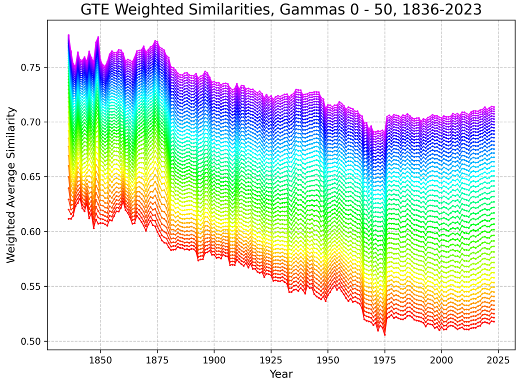




- **Addresses compositional concern:** Decline not driven by shifts across technology fields — spreading out occurs *within* established classes

Backup: Similarity at Different Spatial Scales

Robustness



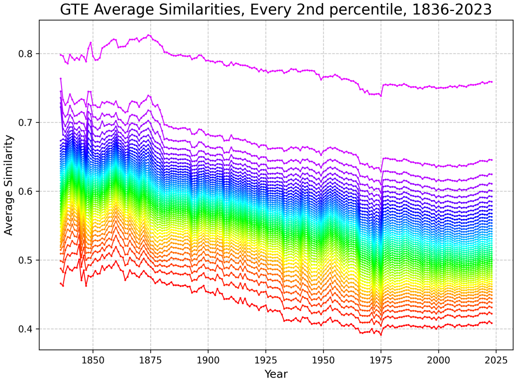
Weighted average: $\equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_{j\neq i}(1-d_{ij})e^{-\gamma d_{ij}}}{\sum_{j\neq i} e^{-\gamma d_{ij}}}$

where γ from 0 (global) to 50 (local)

- **Key finding:** Similar declining trends across all spatial scales
- Model predictions concern averages — important to verify pattern holds across distribution
- Post-2000 arrest slightly stronger at local scales (consistent with entity correction)

Backup: Similarity at Different Quantiles

Robustness

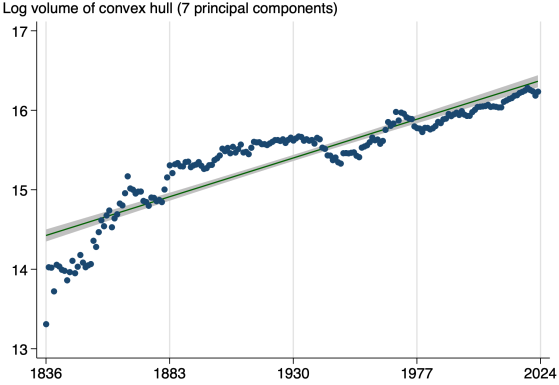


Similarity at Different Quantiles:

- 50 quantiles of pairwise similarity in each year
- Secular decline is robust across all quantiles
- Post-2000 increase in similarity is slightly faster for higher quantiles

Corroboration: Expanding Convex Hull

Robustness   R&D regression



Is Idea Space Expanding?

Test:

- 1024 GTE dimensions to 7 PC
- Measure volume of convex hull

Result:

- +0.5%/yr (6 PC: +0.4%/yr)
- Likely under-estimate due to dimensionality reduction

Log TFP Specification

Return

Why linear in log TFP? $A_i(h) = Q_i - \tau h$

Standard in spatial competition (Salop 1979):

- Idea consumers have preferences linear in quality net of distance costs
- $A_i(h)$ interpreted as log TFP $\Rightarrow$ firms care about *proportional* productivity gains

Microfoundation: Each downstream firm has one unit of fixed input ℓ and produces:

$$y = e^A \cdot \ell$$

With output price = 1 and $\ell = 1$, profit is $\pi = e^A$. Willingness to pay for technology delivering incremental log TFP A (relative to baseline $e^0 = 1$):

$$WTP = e^A - 1 \approx A \quad \text{(first-order Taylor approximation)}$$

Accuracy: For annual TFP increments ($A \approx 0.015$/year), approximation error $< 0.01\%$

Advantage: Predictions directly comparable to empirical TFP elasticities (Bloom et al. 2013) and growth accounting (Bloom et al. 2020)

Backup: Model Equations

◂ Equilibrium ◂ Growth rates

Fixed cost (burden of knowledge):	$f(H) = \phi H^{\alpha}$
R&D cost:	$c(q_i) = \frac{1}{2}\gamma q_i^{1+\eta}$
Realized quality (with spillovers):	$Q_i = q_i + \frac{1}{2}\beta\left(1 - \frac{d}{\lambda}\right)(q_{i-1} + q_{i+1})$
Equilibrium pricing and quality:	$p^* = \tau d, \qquad q^* = \frac{d}{\gamma}$
Equilibrium spacing:	$d^*(H) = \sqrt{\frac{\phi H}{\tau - \frac{1}{2\gamma}}}$
Equilibrium entry:	$n^* = \frac{H}{d^*}$
Equilibrium revenue:	$R^* = p^* \cdot d^* = \tau d^2$
Growth rates (baseline $\alpha = 1, \eta = 1$):	
Spacing, quality, entry:	$g_d = g_q = g_n = \frac{1}{2} g_H$
Fixed cost:	$g_f = g_H$
Aggregate R&D:	$g_{R\&D} = \frac{3}{2} g_H$

Backup: Equilibrium Existence Conditions

Return

Spreading-out condition: Proposition

$$\tau\gamma > \frac{1}{2}$$

Marginal revenue of expanding territory exceeds marginal cost.

Second-order conditions:

- Pricing: $\partial^2 R/\partial p^2 < 0$ (satisfied)
- Quality: $\partial^2 \pi/\partial q^2 = -\gamma < 0$ (satisfied)
- No spatial deviation (verified in paper)

Additional conditions:

- Spillover reach: $d < \lambda$ (spillovers active)
- Full coverage: All downstream firms adopt some technology

Equilibrium: Pricing and Quality

Symmetric equilibrium: n inventors, equal spacing $d = H/n$, identical (p, q)

Return

1. Pricing. Boundary firm at $\tilde{h}$ indifferent between inventor i and neighbor:

$$Q - p_i - \tau\tilde{h} = Q - p - \tau(d - \tilde{h}) \quad \Rightarrow \quad \tilde{h} = \tfrac{d}{2} + \tfrac{p - p_i}{2\tau}$$

Revenue $R_i = 2p_i\tilde{h}$; FOC $\partial R_i / \partial p_i = 0$ yields: $\boxed{p^* = \tau d}$

2. Quality. MR $=$ MC for quality investment ($\partial Q_i / \partial q_i = 1$, $\partial \tilde{h} / \partial Q_i = 1/2\tau$):

$$\frac{p_i}{\tau} = \gamma q_i \quad \Rightarrow \quad \boxed{q^* = \frac{d}{\gamma}} \quad (\eta = 1)$$

3. Zero-profit condition. Substitute $R = \tau d^2$, $c(q^*) = \frac{d^2}{2\gamma}$:

$$\tau d^2 - \frac{d^2}{2\gamma} - f = 0 \quad \Rightarrow \quad \boxed{d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}}$$

Backup: Free Entry Determines Equilibrium Spacing and Inventions

◂ Return

Zero-profit condition:

$$\underbrace{\tau d^2}_{\text{Revenue}} - \underbrace{\frac{d^2}{2\gamma}}_{\text{R\&D cost}} - \underbrace{f}_{\text{Entry cost}} = 0$$

Solving for equilibrium spacing and number of inventions ($n = H/d$):

$$\boxed{d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}} \quad \Rightarrow \quad \boxed{n^* = H\sqrt{\frac{\tau - \frac{1}{2\gamma}}{f}}}$$

Symmetric equilibrium p^*, q^*, d^*, n^* in terms of costs τ, γ, f, and market size H

Backup: Spreading Out — Proposition

Return

Proposition (Spreading Out)

For $\tau\gamma > \frac{1}{2}$, equilibrium spacing increases with opportunity space: $\dfrac{dd^*}{dH} > 0$.
Inventions become **less similar** over time.

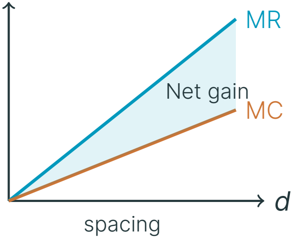
Adaptation costs must create sufficient pricing power

Intuition

Backup: Spreading Out — Why Is It Profitable?

Proposition

Why is spreading out profitable?



Marginal revenue of expanding territory:

- Revenue $R = \tau d^2 \quad \Rightarrow \quad \mathrm{MR} = 2\tau d$

Marginal cost of expanding territory:

- Need higher quality: $q = d/\gamma$
- $\mathrm{MC} = d/\gamma$

Spreading profitable when:

$$\mathrm{MR} > \mathrm{MC} \quad \Rightarrow \quad \boxed{\tau\gamma > \frac{1}{2}}$$

Adaptation costs must create sufficient pricing power

Backup: Comparative Statics Derivations

Equilibrium Literature Evidence

Spreading out: From zero-profit condition $d^2(\tau - \frac{1}{2\gamma}) = \phi H$:

$$\frac{dd}{dH} = \frac{\phi}{2d(\tau - \frac{1}{2\gamma})} = \frac{\phi}{dR/dd - dc/dd} > 0$$

Rising quality and prices:

$$\frac{dq}{dH} = \frac{1}{\gamma}\frac{dd}{dH} > 0, \qquad \frac{dp}{dH} = \tau\frac{dd}{dH} > 0$$

Rising entry:

$$\frac{dn}{dH} = \frac{1}{d} - \frac{H}{d^2}\frac{dd}{dH} > 0 \text{ under spreading-out condition}$$

Declining R&D productivity: See Force decomposition

Backup: Robustness to Entry Cost Curvature

Return

With $f(H) = \phi H^{\alpha}$, equilibrium spacing satisfies $d \propto H^{\alpha/2}$:

Condition	Entry Growth	Prediction
$0 < \alpha < 2$	$g_n = (1 - \frac{\alpha}{2})g_H > 0$	Entry grows ✓
$\alpha = 2$	$g_n = 0$	Entry stagnates
$\alpha > 2$	$g_n < 0$	Entry declines ×

Main results robust for $\alpha < 2$:

- Spreading out: $g_d = \frac{\alpha}{2} g_H > 0$ for any $\alpha > 0$
- Declining R&D productivity: Holds throughout range
- Higher $\alpha \rightarrow$ faster spreading, but lower spatial share of productivity decline

Counterfactual boundary: Patent counts grow over time, ruling out $\alpha \geq 2$

Backup: How Entry Costs Shape Equilibrium

Return

As idea space H grows, what happens to d^*, n^*, q^*, p^*?

Scenario	d^*, q^*, p^*	**Varieties n^***
1. f constant	unchanged	↑ (linear in H)
2. $f(H)$ decreasing	↓ (clustering!)	↑↑ (faster)
3. $f(H)$ increasing (KPF baseline)	↑ (spreading)	↑ (grows with H)
4. $f(H)$ increasing rapidly	↑↑ rapidly	↓ (fewer)

Only rising H, f is consistent with spreading out and expanding varieties

Backup: A Knowledge Production Function

Return

Evidence of expansion:

- Patent filings: ≈500/yr (1840s) → 350,000/yr (2020s)
- Each discovery opens **adjacent research directions** Weitzman 1998; Scotchmer 1991
- Electricity → electronics → computing → AI...

Formalizing the intuition:

$$\dot{H} = \delta \cdot n \cdot \underbrace{(Q - \tau d/4)}_{\text{net quality contribution}}$$

- δ: efficiency of knowledge in opening new directions
- n inventions, each contributing to frontier expansion

Key features:

- Uses *social* quality Q (incl. spillovers), not private effort q
- Net of adaptation costs—only "usable" quality opens new directions
- Standard supply-side KPF (Romer 1990; Jones 1995; Kortum 1997)

Backup: TFP and R&D Growth Equations

◂ Growth rates ◂ TFP regression ◂ R&D regression ◂ Decomposition

TFP growth:

$$g_{TFP} = \underbrace{g_q\left(1+\beta-\frac{\beta d}{\lambda}\right)}_{\text{Quality (with spillovers)}} - \underbrace{\frac{\beta q}{\lambda} g_d}_{\text{Spillover attenuation}} - \underbrace{\frac{\tau}{4} g_d}_{\text{Adaptation drag}}$$

R&D growth:

$$g_{R\&D} = \underbrace{g_n}_{\text{Entry}} + \underbrace{\theta \cdot g_q}_{\text{Quality scaling}} + \underbrace{\theta \cdot g_q}_{\text{Fishing out}} + \underbrace{(1-\theta) g_f}_{\text{Burden of knowledge}}$$

where θ = variable cost share; α = 1, η = 1

Backup: Growth Equations with General Cost Curvatures

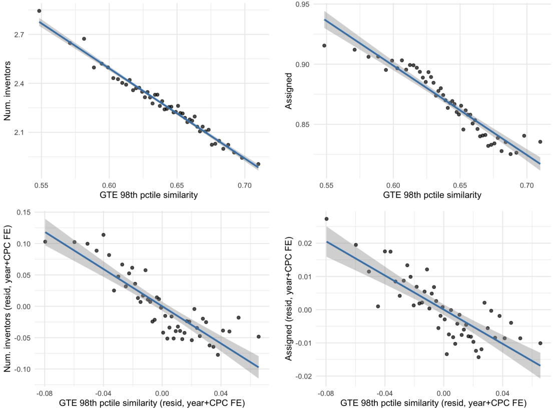
◂ Growth Rates ◂ Decomposition

Component	**General**	**Baseline ($\alpha = 1, \eta = 1$)**
Entry cost	$f(H) = \phi H^{\alpha}$	ϕH
R&D cost	$c(q) = \frac{1}{2}\gamma q^{1+\eta}$	$\frac{1}{2}\gamma q^2$
Spacing growth	$g_d = \frac{\alpha}{2} g_H$	$\frac{1}{2} g_H$
Quality growth	$g_q = g_d$	$\frac{1}{2} g_H$
Entry growth	$g_n = (1 - \frac{\alpha}{2}) g_H$	$\frac{1}{2} g_H$
Fixed cost growth	$g_f = \alpha g_H$	g_H

R&D growth equation:

$$g_{R\&D} = \underbrace{(1 - \tfrac{\alpha}{2}) g_H}_{\text{Entry}} + \underbrace{\tfrac{\theta\alpha}{2} g_H}_{\text{Quality scaling}} + \underbrace{\tfrac{\theta\eta\alpha}{2} g_H}_{\text{Fishing out}} + \underbrace{(1 - \theta)\alpha g_H}_{\text{Burden of knowledge}}$$

Backup: Spacing Predicts Investment and Quality



◂ Return

Model: $q^* = d/\gamma$ — more isolated inventors invest more in quality

4 q proxies: Co-inventors, assigned patents, 5-yr citations, KPSS patent value

Confirmed in cross-section and within technology classes over time

▸ Regression results

Spacing and Quality: Regression Results

Return

	Co-Inventors	Firm Assignment	Citations (5-yr)	KPSS Value
Panel A: Cross-sectional, mean similarity				
GTE	−2.545***	−0.659***	6.347	−22.093**
	(0.679)	(0.180)	(4.480)	(9.036)
PaECTER	−4.928***	−1.723***	−36.681***	−122.904***
	(1.002)	(0.424)	(8.383)	(46.101)
Panel B: Cross-sectional, 98th percentile similarity				
GTE	−1.477**	−0.253*	4.788	−32.531***
	(0.670)	(0.131)	(3.220)	(7.022)
PaECTER	−1.067	−1.174***	−19.404***	−84.301*
	(1.592)	(0.431)	(6.443)	(43.556)
Observations	219,772	219,772	219,772	3,274,108
Fixed Effects	Year + CPC Class			
Panel C: Time-series, Δ mean similarity, CUSP (1836–2023)				
GTE	−0.011*	—	—	—
	(0.005)			
PaECTER	−0.004**	—	—	—
	(0.001)			
Observations	15,813			
Fixed Effects	Year + CPC Class			

Backup: Validation of Spillover Elasticity

Return

Quasi-experimental estimates of R&D spillovers Bloom et al. 2013, Lucking et al. 2019

- Firm-level TFP elasticity to shocks to spillover pool
- SPILLTECH ≡ sum of neighbors' R&D, weighted by idea distance
- IV: State R&D tax credit shocks

Ingredients:

- Bloom et al. 2013 (1981-2001): $\hat{\beta} =$ **0.206**
- Standard deviation log(SPILLTECH) = **1.04**
- Avg change in similarity (Bloom sample period): **-0.007σ/yr**

Implied TFP drag from average rate of spreading out:

- -0.007 × 1.04 × 0.206 = **-0.15%/yr** ✓

Backup: Forces Reducing Research Productivity

◂ Return

Forces reducing g_{TFP}:

$$g_{TFP} = \frac{dA}{dH} \cdot \dot{H} = \left(\underbrace{\frac{dq}{dH}\left[1+\beta\left(1-\frac{d}{\lambda}\right)\right]}_{\textbf{(1) Spillover attenuation: mult.}\downarrow} - \underbrace{\frac{\beta q}{\lambda}\frac{dd}{dH}}_{\substack{\textbf{(2) Spillover}\\ \textbf{base: } q\uparrow}} - \underbrace{\frac{\tau}{4}\frac{dd}{dH}}_{\text{Adapt. drag}} \right) \cdot \underbrace{\left[g_H^* + \frac{\delta\beta}{\gamma}\left(1-\frac{d}{\lambda}\right)\right]}_{\substack{\textbf{(3) Frontier}\\ \textbf{spillovers}\downarrow \text{ as } d\uparrow}} \cdot H$$

Forces increasing Agg R&D:

$$\frac{d(\text{Agg R\&D})}{dH} = \underbrace{\frac{dn}{dH} \cdot [c(q)+f(H)]}_{\textbf{(6) Entry expansion}} + \underbrace{n \cdot c'(q)}_{\textbf{(4) Fishing out}} \cdot \underbrace{\frac{dq}{dH}}_{\textbf{(7) Quality scaling}} + \underbrace{n \cdot f'(H)}_{\textbf{(5) Burden of knowledge}}$$

Spatial Forces: 1, 2, 3, 6, 7 — Traditional: 4, 5

Backup: TFP Estimation Strategy

1. Theory

Derivation | Return

$$g_{TFP} = \underbrace{\frac{dq}{dt}\left[1+\beta\left(1-\frac{d}{\lambda}\right)\right]}_{\text{local ideas + spillovers}} - \underbrace{\left(\frac{\beta q}{\lambda}+\frac{\tau}{4}\right)\frac{dd}{dt}}_{\text{spreading losses}}$$

With $q^* = d/\gamma$: $$g_{TFP} = \underbrace{\kappa_1}_{\text{net quality effect}} \frac{dd}{dt} - \underbrace{\kappa_2}_{>0} d\,\frac{dd}{dt}$$

2. Regression

Table

$$\Delta \log(\mathrm{TFP})_t = b_0 + b_1(-\Delta \mathrm{Sim}_t) + b_2(-\Delta \mathrm{Sim}_t)(-\mathrm{Sim}_{t-1}) + b_3 t + \epsilon_t$$

b_1: quality gains net of adaptation costs

$b_2 < 0$: spreading weakens spillovers more at greater distance

3. Quantification

At observed spreading and 1991 spacing

$\widehat{\mathrm{drag}} =$ **−0.16** pp/yr

Explains **7%** of the 1948–2015 TFP-growth slowdown

Backup: R&D Estimation Strategy

Derivation Return

1. Aggregate R&D accounting

$$g_{R\&D} = \underbrace{g_n}_{\text{inventions}} + \underbrace{\theta(1+\eta)g_q}_{\text{variable quality costs}} + \underbrace{(1-\theta)g_f}_{\text{fixed costs}}$$

2. Equilibrium relationships

$$n = H/d, \qquad q = d/\gamma, \qquad f(H) \propto H^{\alpha}$$

$$\Downarrow$$

$$g_n = g_H - g_d, \qquad g_q = g_d, \qquad g_f = \alpha g_H$$

3. Substitute into aggregate R&D growth

$$g_{R\&D} = \underbrace{[1+\alpha(1-\theta)]g_H}_{a_0} + \underbrace{[\theta(1+\eta)-1]g_d}_{a_1 g_d}$$

Empirical equation $(g_d \approx -\Delta\mathrm{Sim}_t)$: $g_{R\&D,t} = a_0 + a_1(-\Delta\mathrm{Sim}_t) + \epsilon_t$

Regression table

Slope identifies the variable-cost share

$$a_1 = \theta(1+\eta) - 1 \quad \Rightarrow \quad \boxed{\theta = \frac{a_1+1}{1+\eta}}$$

Intercept identifies frontier growth

$$a_0 = [1+\alpha(1-\theta)]g_H \quad \Rightarrow \quad \boxed{g_H = \frac{a_0}{1+\alpha(1-\theta)}}$$

Backup: Calibrating the Model

Return

What we assume (baseline):

Parameter	Value	Source/Interpretation
θ (variable cost share)	0.69	NSF survey; regression estimate (0.72)
g_H (idea space growth)	2.7%/yr	Two methods: $4.0\% = g_{R\&D} = 3/2 \cdot g_H$; R&D regression constant a_0 ($\alpha = 1$)
α (entry cost curvature)	1.0	Entry costs scale linearly with H
η (R&D cost curvature)	1.0	Quadratic R&D costs

Backup: Detailed Spatial Decomposition

This is a **model-conditioned decomposition**, disciplined by external checks above.

Calibration: $\theta = 0.69$ (NSF) $g_H = 2.7\%/\text{yr}$ (Model identity $g_{R\&D} = \frac{3}{2} g_H$) $\alpha = 1$, $\eta = 1$ (baseline)

Backup: Detailed Spatial Decomposition

This is a **model-conditioned decomposition**, disciplined by external checks above.

Calibration: $\theta = 0.69$ (NSF) $g_H = 2.7\%/\text{yr}$ (Model identity $g_{R\&D} = \frac{3}{2} g_H$) $\alpha = 1$, $\eta = 1$ (baseline)

Component	Contribution
TFP deceleration	–1.6%/yr
Spatial drag worsened	–0.11%/yr
Unmodeled factors	–1.49%/yr
R&D growth	+4.0%/yr
Total R&D productivity decline, 1948–2015	–5.6%/yr

Spatial drag: Knowledge spillovers fade

▸ Robustness ▸ Growth equations ◂ Return

Backup: Detailed Spatial Decomposition

This is a **model-conditioned decomposition**, disciplined by external checks above.

Calibration: $\theta = 0.69$ (NSF) $g_H = 2.7\%$/yr (Model identity $g_{R\&D} = \frac{3}{2} g_H$) $\alpha = 1$, $\eta = 1$ (baseline)

Component	Contribution
TFP deceleration	–1.6%/yr
Spatial drag worsened	–0.11%**/yr**
Unmodeled factors	–1.49%/yr
R&D growth	+4.0%/yr
Entry expansion $(1-\frac{\alpha}{2})g_H$	+1.33%**/yr**
Quality scaling $(\theta \frac{\alpha}{2} g_H)$	+0.92%**/yr**
Fishing out $(\theta \eta \frac{\alpha}{2} g_H)$	+0.92%/yr
Burden of knowledge $(1-\theta)(\alpha g_H)$	+0.83%/yr
Total R&D productivity decline, 1948–2015	–5.6%/yr

Entry expansion: More inventions on more territory

Quality scaling: Each inventor builds bigger, raising cost per unit TFP

▸ Robustness ▸ Growth equations ◂ Return

Backup: Detailed Spatial Decomposition

This is a **model-conditioned decomposition**, disciplined by external checks above.

Calibration: $\theta = 0.69$ (NSF) $g_H = 2.7\%$/yr (Model identity $g_{R\&D} = \frac{3}{2} g_H$) $\alpha = 1$, $\eta = 1$ (baseline)

Component	Contribution
TFP deceleration	−1.6%/yr
Spatial drag worsened	−0.11%/yr
Unmodeled factors	−1.49%/yr
R&D growth	+4.0%/yr
Entry expansion $(1-\frac{\alpha}{2})g_H$	+1.33%/yr
Quality scaling $(\theta \frac{\alpha}{2} g_H)$	+0.92%/yr
Fishing out $(\theta \eta \frac{\alpha}{2} g_H)$	+0.92%/yr
Burden of knowledge $(1-\theta)(\alpha g_H)$	+0.83%/yr
Total R&D productivity decline, 1948–2015	−5.6%/yr
Spatial contribution (42%)	−2.36%/yr
Non-spatial contribution (58%)	−3.24%/yr

Spatial forces account for 42% of the long-run productivity decline

Sensitivity: **credible range 42–58%**; baseline (42%) is conservative; better-identified calibrations higher

▸ Robustness ▸ Growth equations ◂ Return

Backup: TFP Regression Results

Return

	Annual		3-Year	5-Year
$b_1 : -1 \times \Delta\text{Sim}$	-0.169^{***}	-0.171^{***}	-0.278^{***}	-0.269^{***}
	(0.057)	(0.083)	(0.095)	(0.098)
$b_2 : (-1 \times \Delta\text{Sim}) \times (-1 \times \text{Sim}_{t-1})$	—	-0.015	-0.408	-0.571^{*}
		(0.342)	(0.320)	(0.312)
Implied TFP drag from spreading out ($\overline{\Delta Sim}$, %/yr):				
1948 (Sim $= 0.35$)	-0.08	-0.08	-0.07	-0.04
1991 (Sim $= 0$, baseline)	-0.08	-0.09	-0.14	-0.16

$\hat{b}_1 < 0$: adaptation drag > quality investment $\hat{b}_2 < 0$: spillover attenuation confirmed ✓

Data: TFP and Real R&D, 1948–2015 (Bloom et al., 2020)

Backup: Implied $\theta = 0.72$ and $g_H = 2.7\%$/yr

◂ Return

	Annual	3-Year	5-Year
a_1: $-1 \times \Delta$Sim	0.165	**0.448****	**0.438***
	(0.177)	(0.219)	(0.244)
a_0: Constant	**0.034*****	**0.102*****	**0.173*****
	(0.006)	(0.013)	(0.018)
Implied θ (variable cost share)	0.58	0.72	**0.72**
Implied g_H (idea space growth)	2.4%/yr	2.7%/yr	**2.7%/yr**

- $\theta = 72\%$ aligns with NSF survey data (labor $= 69\%$ of R&D) ✓
- $g_H = 2.7\%$/yr consistent w/ embedding volume growth & $g_{R\&D} = 3/2 \cdot g_H$ ✓ ▸
- Model predicts $g_d / g_{R\&D} = 1/3$; in data, $-\Delta\text{Sim}/g_{R\&D} \approx 0.3$ ✓

Backup: Spatial Share Higher for Better-Identified Calibrations

◂ Return

Dominant uncertainty: α (entry cost curvature)

Evidence pins $\alpha \approx 0.67$ unique inventor growth (2.0%/yr)

$\eta = 0.625$ from quasi-experiment Guceri-Liu 2016

Baseline is conservative: Credible range 42–58%

	Spatial share (%)	
α	$\eta = 1$	$\eta = 0.625$
0.50	55	**58**
0.75	48	51
1.00	**42**	46
1.25	37	41
1.50	33	37