Rising Idea Complexity and the Organization of Innovation Teams

Stefano Cravero - April 29th, 2026

University of Pennsylvania

Motivation and Research Question

- The nature of innovation has been changing in a fundamental way.
  - Solo-inventor's patents: from $\approx 60\%$ in 1976 to less than 30% today. [USPTO]
  - Teams combine knowledge from multiple different domains: *e.g., 2024 Chemistry Nobel.*
- **This project**: ideas are growing increasingly complex.
  - As frontier advances, innovation requires combining expertise from different fields.

$\Longrightarrow$ Expect individuals of different background to start working together.

- Implications go beyond the organization of innovation teams; *for instance*:
  - Firm dynamics: innovation entry cost $\uparrow$, innovation concentration $\uparrow$

$\Longrightarrow$ *Question*: Can this explain the observed increase in market concentration?

Pre-requisite Questions

- **Measurement**: how do we measure the knowledge of inventors and teams?
- **Mechanism**: why does innovation increasingly rely on teams and diverse skills?

Status Quo of the Project

Empirics:

- Construct a measure of team-level **skill heterogeneity** using inventor's job history.
- Show how the average span of teams' skills has increased over time.
- Show that there exists **complementarities** across skills.
  - Team heterogeneity $\downarrow \implies$ Team's and firm's productivities $\downarrow$

Theory (*in progress*):

- Ideas become more complex as they are drawn further down the tail.
- *Why?* Knowledge required for innovation may come from multiple different domains.
  $\implies$ Heterogeneous skill areas become complementary in the innovation process.
  $\implies$ Teams expand to absorb inventors with different skills.

Data Sources for Empirical Analysis

INV-BIO ADIAB:

- Linked inventor biography data 1980-2014.
- Inventor panel containing:
  - **Full labor market histories** over 1980-2014 for all inventors who patented in Germany over 1999-2011.
  - **Complete patent records** to which the individual contributed over 1980-2014.
  - **Employing establishment characteristics** over 1980-2014.
- $\approx$ 150k unique inventors.
- $\approx$ 650k patent files ($\approx$ 71% of all patents granted from 1999 to 2011).
- $\approx$ 150k establishment IDs ($\approx$ 97% matched with their characteristics).

Inventor-level Skill Profiles

Central Question

What skills do individuals posses?

- Proxy skills using inventors' experience in $\mathbf{K = 37}$ 2-digit occupations (*e.g., engineering, physics, computer science*).
- **Inventor skills**: given inventor i in year t, count days $c_{k,i,t}$ worked in occupation k over rolling window (e.g., full job history, last 10 years, etc.). Construct

$$p_{k,i,t} = \frac{c_{k,i,t}}{\omega_i}, \quad \text{where } \omega_i \equiv \sum_{k=1}^{K} c_{k,i,t}.$$

  Let $\vec{p}_{i,t} = (p_{1,i,t}, \ldots, p_{K,i,t})$ be the **inventor-level skill vector**.

- **Example**: inventor i worked 3 years as an engineer and 1 as physicist:

$$\vec{p}_{i,t} = (\underbrace{0.75}_{\text{Engineering}}, \underbrace{0.25}_{\text{Physics}}, 0, \ldots, \underbrace{0}_{\text{Computer Science}})$$

Team-level Skill Profiles

Central Question

What skills do *teams of individuals* posses?

- **Team skills**: given patent p filed in year t with inventor set T_p, let team skills be

$$p_{k,t}^{(\text{team})} = \frac{\sum_{i \in T_p} \omega_i p_{k,i,t}}{\sum_{i \in T_p} \omega_i}.$$

  Let $\vec{p}_t^{(\text{team})} = (p_{1,t}^{(\text{team})}, \ldots, p_{K,t}^{(\text{team})})$ be the *experience-weighted* **team-level skill vector**.

- **Example**: inventor 1 worked 9 years as engineer and inventor 2 worked 1 year as physicist:

$$\vec{p}_t^{(\text{team})} = (\underbrace{0.9}_{\text{Engineering}}, \underbrace{0.1}_{\text{Physics}}, 0, \ldots, \underbrace{0}_{\text{Computer Science}})$$

Are Teams Becoming More Heterogeneus?

Central Question

How do we map team's skill into skill heterogeneity?

- Measure within-team heterogeneity as the dispersion (**entropy**) of the skill-vector:

$$E(\vec{p}) = -\sum_{k=1}^{K} p_k \ln(p_k) \in [0, 1] \qquad \text{with } 0 \ln(0) \equiv 0.$$

- **Extreme cases**:
  - One-skill team (e.g., all inventors are engineers): $E(\vec{p}) = 0$.
  - All skills are equally represented (i.e., $p_k = 1/K$): $E(\vec{p}) = 1$.

Are Teams Becoming More Heterogeneus?

- Team heterogeneity could be increasing for two reasons:
  - Inventors switch occupation more often: $E(\vec{p}_{i,t}) \uparrow \implies E(\vec{p}^{(team)})\uparrow$
  - **My focus**: inventors sort in more heterogeneous teams: $E(\vec{p}^{(team)})\uparrow$
- Use across-individual heterogeneity as

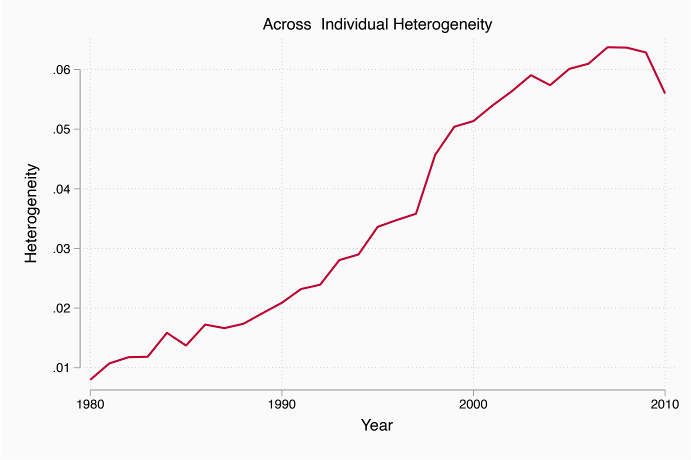
$$H_{p,t} = \frac{1}{\ln(K)}\left[\underbrace{E\left(\vec{p}_t^{(team)}\right)}_{\textbf{Within-Team Heterogeneity}} - \underbrace{\frac{1}{|T_p|}\sum_{i\in T_p} E(\vec{p}_{i,t})}_{\textbf{Average Within-Individual Heterogeneity}} \right] \in [0,1].$$

- **Extreme cases**:
  - All inventors on the team have identical skill background: $H_{p,t} = 0$.
  - All inventors are different from each other *and* all skills are equally represented: $H_{p,t} = 1$.

Cosine Distance

Are Teams Becoming More Heterogeneous?

- Inventors are sorting into teams six times more heterogeneous is 2010 vs 1980.



No weighting | 10-year windows | Current-occupation | All teams | Within-Team Heterogeneity | Within-Individual Heterogeneity (10y)

Decomposing the Rise in Heterogeneity

- The increase in team heterogeneity can be decomposed into three effects:
  1. Reallocation: teams are becoming larger $\implies H\uparrow$ mechanically
  2. Inventor pool-drift: the universe of inventors is becoming more heterogeneous $\implies H\uparrow$
  3. Excess heterogeneity: conditional on team size, inventors collaborate with increasingly diverse partners $\implies H\uparrow$
- Let $w_t(m)$ be the share of patents with team-size m in year t. Then:

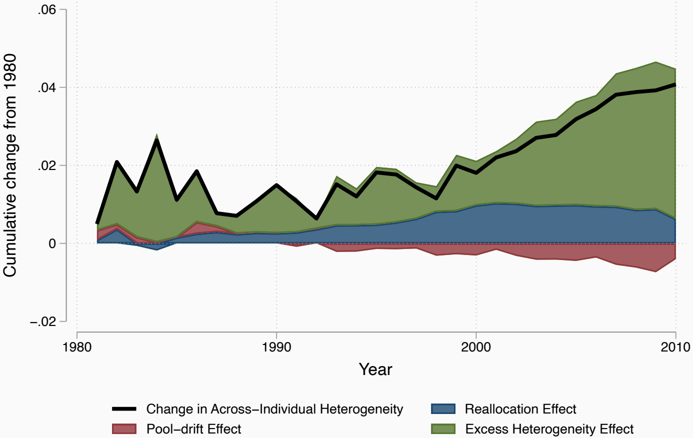
$$\Delta H_t = \sum_m \left(\frac{w_{t-1}(m) + w_t(m)}{2}\right) \overbrace{\Big(\underbrace{\Delta g_t(m)}_{\text{Pool-drift Effect}} + \underbrace{\Delta \mathcal{E}_t(m)}_{\text{Excess Heterogeneity Effect}} \Big)}^{=\Delta H_t(m)}$$

$$+ \underbrace{\sum_m \Delta w_t(m) \left(\frac{H_{t-1}(m) + H_t(m)}{2}\right)}_{\text{Reallocation Effect}}$$

Decomposing the Rise in Heterogeneity

- **Conditionally** on teams ($n \geqslant 2$), 80% of increase comes from teams becoming more diverse within size (excess heterogeneity).

Unconditional Decomposition



How Do Skill Complementarities Affect Firm Outcomes?

Central Question

How does skill composition affect team's and firm's productivity?

- **Dynamic team defintion**: let inventors $\mathcal{P} \equiv \{i, j, \dots\}$ patent together at time t. Then:
  - Team formation at $t - h$: all $i \in \mathcal{P}$ recorded in same firm for the first time.
  - Inventor i leaves team $\mathcal{P}$ at $t + k$ when it moves to a different firm.

Event study: loss of unique skill ("*treatment*") vs loss of redundant skill ("*control*").

- Empirical specification:

$$y_{f,\tau} = \underbrace{\alpha_e + \lambda_{year}}_{\text{Firm + Year FE}} + \sum_{\tau \neq -1} \underbrace{\theta_\tau D_{f,\tau}}_{\text{Control Group}} + \sum_{\tau \neq -1} \underbrace{\beta_\tau \left(D_{f,\tau} \times \text{Unique Leaver}_f\right)}_{\text{Treatment Group}} + \varepsilon_{f,\tau}$$

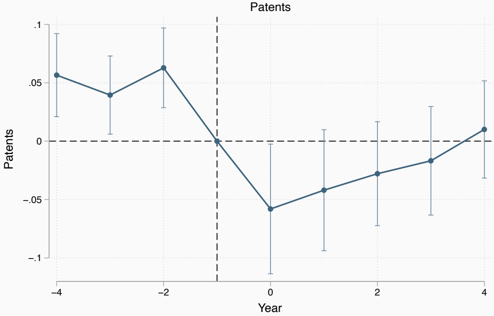
  - Outcomes $y_{f,\tau}$: patent counts and citation counts.

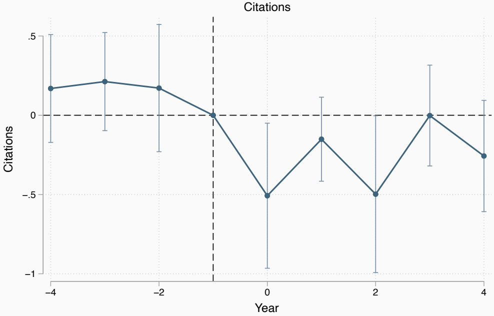
How Do Skill Complementarities Affect Firm Outcomes?

Team-level outcomes:

Firm Employment

- Patents and citations decline significantly and persistently after the team loses an inventor supplying a **unique** skill.





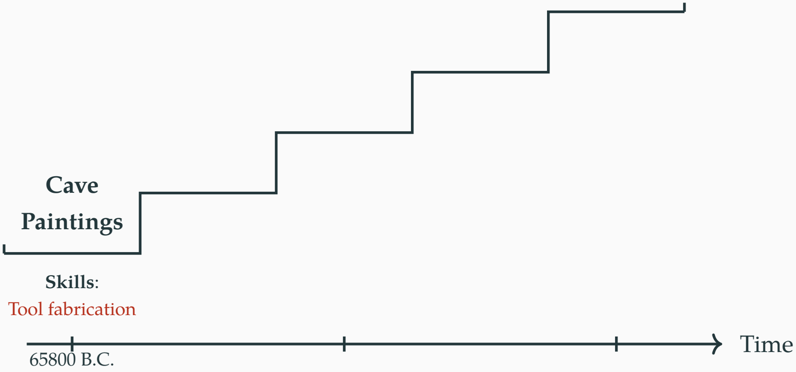
Take-Aways from Empirical Analysis

1. Innovation teams are growing increasingly skill-diverse over time.
2. This trend is largely driven by increased sorting of inventors of different backgrounds into the same innovation team.
3. Heterogeneous skills appear to be complementary in the innovation process.

Idea of a Theory - An Example

History of communication technologies:

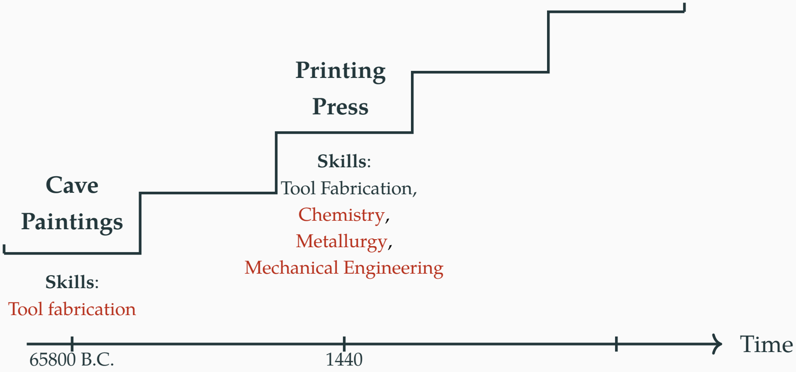
- **Step 0**: Cave Paintings
  - Required knowing using flints to engrave and which materials to use for colors.



Idea of a Theory - An Example

History of communication technologies:

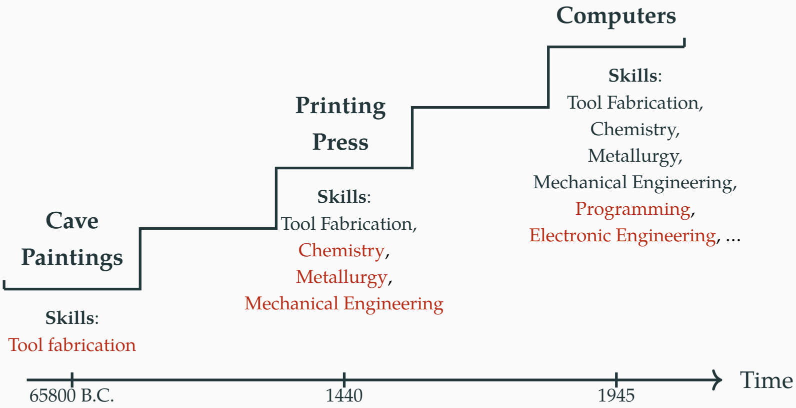
- **Step n**: Printing Press
  - Required chemical experimentation, knowledge of metallurgy, etc.



Idea of a Theory - An Example

History of communication technologies:

- **Last Step** (so far): Computers
  - Required programming skills, electronic engineering, ...



Idea of a Theory - Necessary Ingredients

Central Question

Why do firms want to assemble increasingly **skill-heterogeneous** teams over time?

- Each productive process (**product line**) is a *ladder*.
- An **innovation** is represented by the successful climbing of one step of the ladder.
- Climbing each step requires mastering new knowledge.
  - The innovation ladder is a collection of **knowledge blocks**.
- Each knowledge block belongs to a *knowledge domain* (**skill area**).
  - Steps $\uparrow \implies$ Number of knowldge blocks $\uparrow \implies$ Number of skill areas $\uparrow$
  - As the project advances, the knowledge that underpins it becomes more **complex**.
- Knowldge complexity $\uparrow \implies$ firms assemble inventors who can master *different* skills.
  - Average team's innovation heterogeneity $\uparrow$ as the productivity frontier advances.

Two-Skill Benchmark - Setting

- Firm producing product p of current technological development level z.
- A new innovative idea:
  - arrives at rate $A(z)$
  - yields benefit $\Delta V(z)$ if completed
  - requires $m(z)$ knowledge blocks: first is *core block*, remaining are *auxiliary blocks*:

$$R_p = \{ \underbrace{r_{p1}}_{\text{Core Block}}, \underbrace{r_{p2}, \ldots, r_{pm}}_{\text{Auxiliary Blocks}} \}$$

- Knowledge blocks belong to two skill areas, $s \in \{1, 2\}$: $r_{pj} \in \{1, 2\}\ \forall j \leqslant m$.
  - Auxiliary block $m(z) \geqslant 1$ is in the same domain of core block with probability $\chi \in (0, 1)$.
- Blocks inside the team's skill set are handled with efficiency 1, those outside with efficiency $\eta \in [0, 1)$.

Two-Skill Benchmark - Innovation Process

- Core skill $s(r_{p1}) \in \{0, 1\}$ is required in the team to innovate.
- Firm can opt for:
  - Narrow team (one skill): each knowledge block r is completed with expected efficiency:

$$\underbrace{\chi}_{\text{Skill \textbf{inside} the team, efficiency 1}} + \underbrace{(1-\chi)\eta}_{\text{Skill \textbf{outside} the team, reduced efficiency } \eta < 1}$$

    Each of the $m(z)$ blocks must be completed *simultaneously* and *independently*. The total success arrival rate is:

$$Q_n(z) = [\chi + (1-\chi)\eta]^{m(z)-1}$$

  - Broad team (both skills): every knowledge block is completed with efficiency 1. The total success arrival rate is:

$$Q_b(z) = 1 \geqslant Q_n(z)$$

Two-Skill Benchmark - Choice of Team Heterogeneity

- Firm maximizes expected profit from the innovation

$$\max \quad \{\underbrace{A(z)Q_n(z)\Delta V(z)}_{\text{Narrow Team}}, \quad \underbrace{A(z)Q_b(z)\Delta V(z) - \kappa}_{\text{Broad Team}}\}$$

where $\kappa > 0$ is a span-of-control cost required to operate a broad team.

- The relative efficiency of narrow teams is decreasing in the project complexity $m(z)$:

$$\frac{Q_n(z)}{Q_b(z)} = [\underbrace{\chi + (1-\chi)\eta}_{<1}]^{m(z)-1}$$

- $\exists$ a development threshold z^* such that the firm chooses a broad team iff $z \geqslant z^*$.
- The framework can be extended to arbitrary number of skills $M \in \mathbb{N}$.

Continuous-Skill Model

- Continuum of skill areas: $s \in [0, 1]$.
- Let $A \subseteq [0, 1]$ be the subset of skills belonging to the team.
  - Team's heterogeneity is the measure of active skills: $B = \mu(A)$.
- Each knowledge block $j \in [1, m(z)]$ is completed with success rate

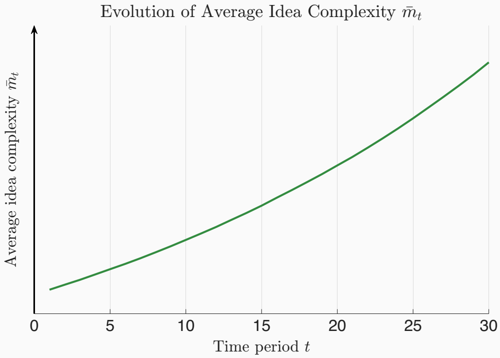
$$Q(B) \equiv \underbrace{\chi + (1-\chi)B}_{\text{Skill } j \in A} + \underbrace{\eta(1-\chi)(1-B)}_{\text{Skill } j \notin A} = \eta + (1-\eta)[\chi + (1-\chi)B]$$

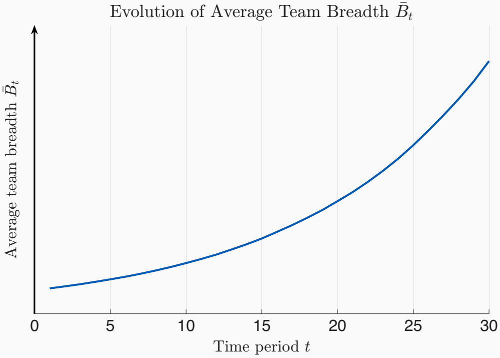
- Total success arrival rate for project of complexity $m(z)$ is $Q(B)^{m(z)}$.
- The firm solves

$$\Pi(z; B) = \max_{B \in [0,1]} \underbrace{A(z)Q(B)^{m(z)}}_{\text{Total Innovation Arrival Rate}} \cdot \Delta V(z) - \underbrace{\Phi(B)}_{\text{Span-of-control Cost}}$$

Continuous-Skill Model - Simulation

- As technological level z increases, so does average idea complexity m.
- Firms adjust by increasing the team's average skill breadth B (heterogeneity).





Conclusions and Next Steps

- Inventors are progressively sorting into more heterogeneous teams.
- Complementarities across heterogeneous skills affect innovation productivity.
- Increasing idea complexity can explain this trend.
  - Further innovation requires knowledge that comes from multiple different domains.
  - Innovation teams expand their expertise to improve their efficiency.
- **Next steps**:
  - Introduce inventors of heterogeneous abilities and expertise in the model.
  - **Q.** How do inventors sort into innovation teams as ideas become more complex?
  - Study **implications** for firm dynamics and mobility in the inventor labor market.

References

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Nobel Prizes Methodology

Back

- Each Nobel Prize *winner* is categorized in a *single* contribution area.
  - **Areas**: Theory, Experiment/Empirical, Methods/Tools, Materials/Engineering, Application/Clinical, Computational/Algorithmic, Other.

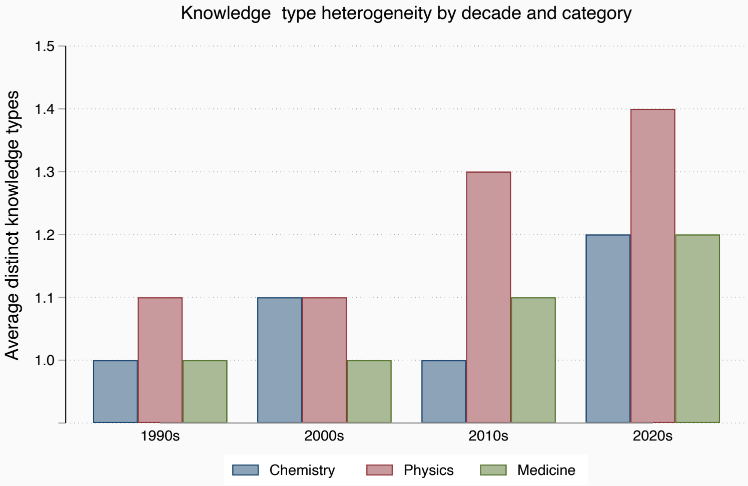
- Categorization is based on the official prize press release.

- **Example**: 2021 Nobel Prize for Physics awarded for the study *"chaotic and apparently random phenomena"*.
  - Syukuro Manabe: *"led the development of physical models of the Earth's climate"* $\Longrightarrow$ **Methods/Tools**
  - Klaus Hasselmann: *"developed methods for identifying specific signals, fingerprints* [...]*"* $\Longrightarrow$ **Methods/Tools**
  - Giorgio Parisi: *"important contributions to the theory of complex systems."* $\Longrightarrow$ **Theory**
  - Two knowledge types underlying the Nobel Prize.

Motivation and Research Question

Back

- Nobel Prizes are increasingly awarded to contributions spanning more than one knowledge field.



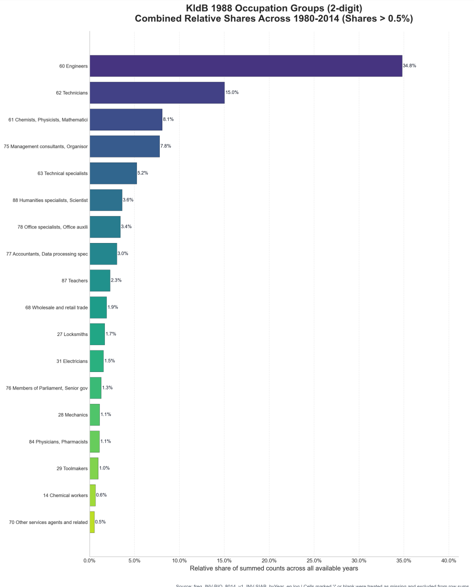
Cosine Distance (1/2)

- Alternative skill heterogeneity measure:
  - **Q.** How **different** are, on average, two inventors belonging to the same team?
- Use **average pairwise cosine distance**:

$$H_p = \frac{2}{|T_p|(|T_p - 1|)} \sum_{i<j \in T_p} \left(1 - \frac{\vec{p}_{i,t} \cdot \vec{p}_{j,t}}{\|\vec{p}_{i,t}\| \, \|\vec{p}_{j,t}\|} \right) \in [0, 2]$$

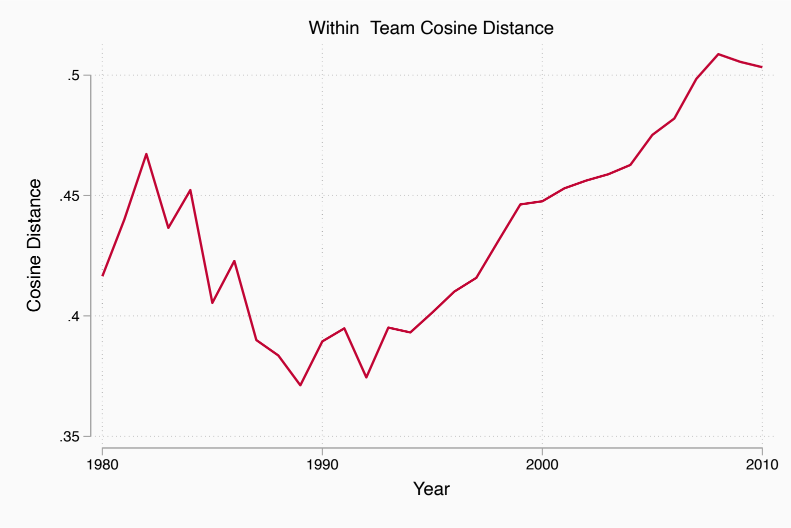
- **Extreme cases**:
  - Perfect positive correlation between inventors' skill vectors: $H_{p,t} = 0$.
  - Perfect negative correlation between inventors' skill vectors: $H_{p,t} = 2$.

Distribution of Individuals across Occupations



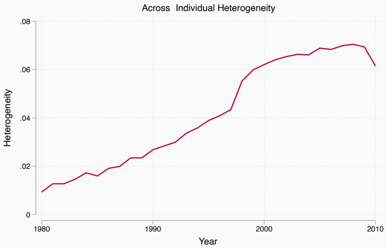
Cosine Distance (2/2)

Back

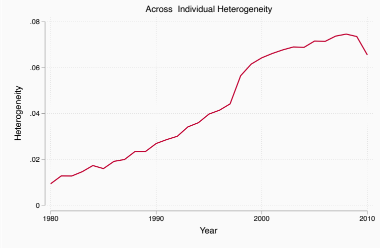


Alternative specifications for across-individual heterogeneity

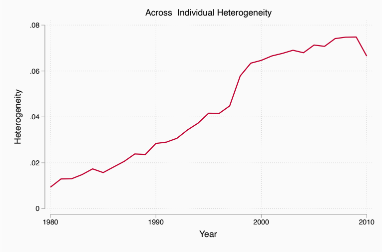
Back



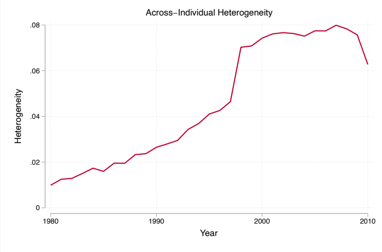
(a) No experience weighting



(b) Using last 10 years of labor history only



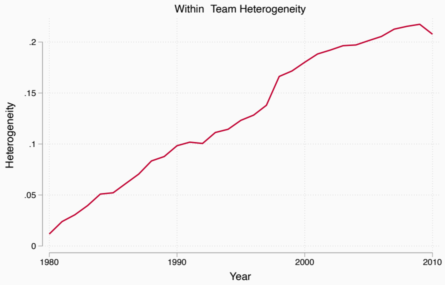
(c) Using current occupation only



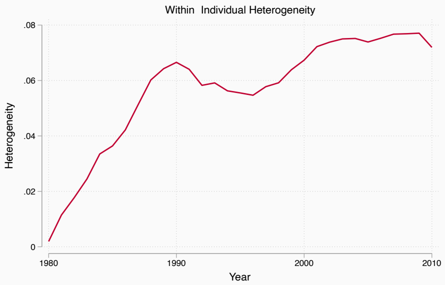
(d) Including Incomplete team patents

Within-Team and Within-Individual Heterogeneity

Back



(a) Within-Team Heterogeneity



(b) Within-Individual Heterogeneity (10-year)

Decomposing the rise in heterogeneity (1/3)

Consider:

- a patent p, filed in year t, with inventor set T_p and of size $m \equiv |T_p|$;
- the across-individual heterogeneity H_p of the patent team filing p.

Define:

- average heterogeneity of teams of size m in year t,

$$H_t(m) \equiv \mathbb{E}_t[H_p \| T_p| = m] ;$$

- empirical distribution of inventor skill profiles in year t, P_t;
- null expected heterogeneity under random matching:

$$g_t(m) \equiv \mathbb{E}[H \mid \text{team of size } m \text{ formed by random draws from } P_t] .$$

Decomposing the rise in heterogeneity (2/3)

- For each team of size m, define excess heterogeneity as $\mathcal{E}_p \equiv H_p - g_t(m)$.
- By construction,

$$H_p = g_t(m) + \mathcal{E}_p \implies H_t(m) = g_t(m) + \underbrace{\mathbb{E}_t[\mathcal{E}_p \| T_p| = m]}_{\mathcal{E}_t(m)}.$$

Olley and Pakes (1996)-like Decomposition

Let $w_t(m)$ be the share of patents with team-size m in year t. Then, the change in within-team heterogeneity can be decomposed as:

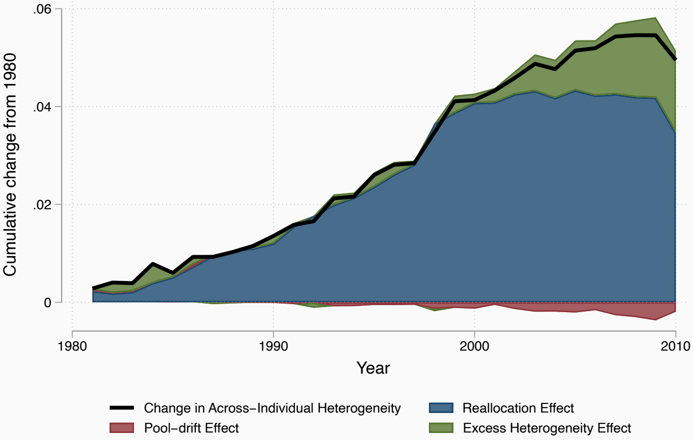
$$\Delta H_t = \underbrace{\sum_m w_{t-1}(m)\, \Delta g_t(m)}_{\textbf{Pool-drift Effect}} + \underbrace{\sum_m w_{t-1}(m)\, \Delta \mathcal{E}_t(m)}_{\textbf{Excess Heterogeneity Effect}} + \underbrace{M_t \cdot \mathrm{Cov}_t^u(H_t(m), \Delta w_t(m))}_{\textbf{Reallocation Effect}},$$

where $\mathrm{Cov}^u(\cdot)$ denotes the *unweighted* covariance.

Decomposing the Rise in Heterogeneity

Back

- **Unconditionally**, larger teams (reallocation) explain 75% of increase in heterogeneity
    - Driven by shift away from solo-inventors.



How Do Skill Complementarities Affect Firm Outcomes?

Back

Firm-level outcomes:

- Firms grow more slowly after the departure of an inventor supplying a unique skill to the team.