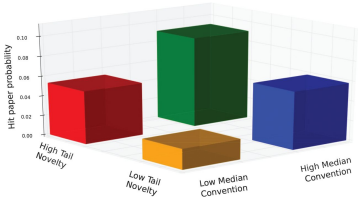
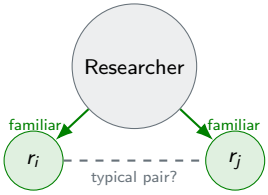
Knowledge Recombination is Shaped by Retrieval Structure



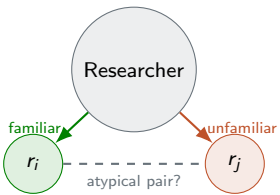
Atypical combinations, *Uzzi et al. (2013)*

Rarer combinations may lead to higher impact.

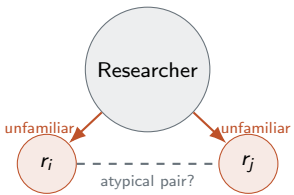
Local search



Pivoting (anchored exploration)



Leaping (unanchored exploration)



Two Dimensions of Knowledge Recombination

Invention (Innovation) is **a process of recombinant search** *(Fleming, 2001)*.

- **Search process**: search/retrieve existing knowledge components.
  - e.g., physicists are more likely to retrieve physics than biology.
  - Conditional probability $\Pr(k_i \mid H_a)$.
- **Combination process**: combine knowledge components into new configurations.
  - e.g., physics and astrophysics are more likely to be combined than physics and biology.
  - Joint probability $\Pr(k_i, k_j)$.

Component unfamiliarity (Local distance, retrieval cost)	Combination atypicality (Global uncommonness, rarity)
How likely is a researcher to reach r_i?	How unusual is the pair (r_i, r_j) in science?
Exploitation / exploration	Typical / atypical
$S(r\|H_a) = -\log \Pr(r\|H_a)$	$A(r_i, r_j) = -\text{npmi}(r_i, r_j)$

Measurement: Trajectory-conditioned Component Unfamiliarity

- **Component unfamiliarity**: how unexpected a knowledge component (cited journal) is given a researcher's prior retrieval trajectory.
- A sequential model $\mathcal{M}$ predicts journal r a researcher is likely to cite in the next paper, conditional on their prior publication history H_a:

$$\Pr(r|H_a) = \mathcal{M}(H_a)$$

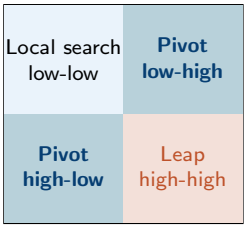
- For each observed cited journal r_i, retrieval surprisal is defined as:

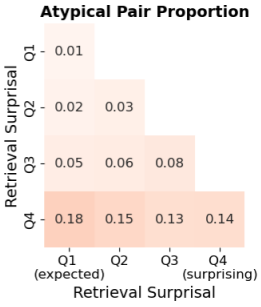
$$S(r_i|H_a) = -\log \Pr(r_i|H_a)$$

- Low value means the component is locally accessible (familiar) to the researcher; high value means the component is distant (unfamiliar) to the researcher.

Results: Anchored Atypicality

- **Pair-level structure**





- Atypicality is highest in asymmetric retrieval profiles, where one component is accessible and the other is distant.

- **Paper-level consequences**

Pair type	Top-5% citation	Disruption index
Anchored atypical	4.367***	−0.003***
Unanchored atypical	2.608***	0.002***
Anchored typical	1.249***	−0.002***
Unanchored typical	0 (baseline)	0 (baseline)

- Anchored atypical pairs are more likely to be highly cited than unanchored atypical pairs.
- Anchored atypical pairs consolidate, while unanchored atypical pairs disrupt.

Takeaway

Contribution

I proposed a trajectory-conditioned component-level measure of **unfamiliarity**: $S(r|H_a)$.

- It captures **how far a single knowledge component is** from a researcher's prior trajectory.
- This makes research trajectories diagnosable:
- When do researchers successfully pivot?
  - (Paper-level view) High exploration breadth $\rightarrow$ lower impact.
  - (Component-level view) Exploration breadth is an aggregate of component-level unfamiliarity.
  - After accounting for anchor share, exploration breadth is positively associated with impact.
  - **Exploration penalty** is not simply about going far, but about going far without anchors.
- How do researchers' careers evolve?
- How do collaborators integrate their different expertise?