

Pivoting, not Leaping

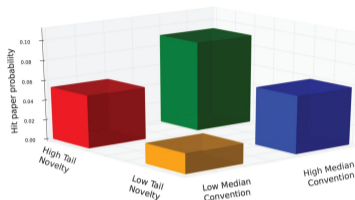
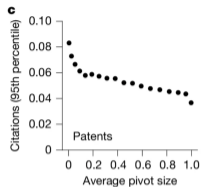
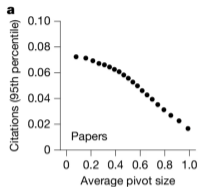
Atypical Knowledge Combinations with Anchored Exploration

Saeyoung Choi, Junghye Lee

Seoul National University

Daejeon Conference on Science and Innovation Policy

Motivation

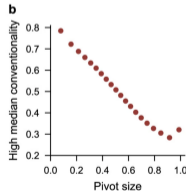
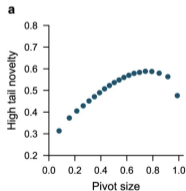


Pivot penalty, Hill et al. (2025)

Broader exploration may lead to lower impact.

Atypical combinations, Uzzi et al. (2013)

Rarer combinations may lead to higher impact.



Exploration and atypicality

Broader exploration is associated with higher atypicality.

Atypical combination benefit

- High-impact science often combines knowledge components that rarely appear together.
- This view treats novelty as distance in the global knowledge space.

Exploration penalty

- Researchers who pivot into unfamiliar domains often lose accumulated intellectual capital.
- This view treats novelty as movement away from a researcher's trajectory.

- If exploration produces atypicality, why does it also carry an impact penalty?
- How can scientists produce valuable atypical combinations without paying the full cost of exploration?
- To interpret this tension, we need to integrate two axes that have been studied separately: **global combination atypicality** and **researcher-specific (local) accessibility**.

Innovation is a recombinant search process that involves

- (1) retrieving existing knowledge components
- (2) and combining them into new configurations.

- **Retrieval process** (*exploration*): search/retrieve existing knowledge components.
 - The same knowledge component can be easily accessible to one but not to another, depending on their prior trajectories and expertise.
 - e.g., **physicists** are more likely to retrieve **physics** than **biology**.
 - Conditional probability $\Pr(k_i | H_a)$.
- **Combination process**: combine knowledge components into new configurations.
 - Some combinations (pairs) are commonly observed, while others rarely appear together.
 - e.g., **physics** and **astrophysics** are more likely to be combined than **physics** and **biology**.
 - Joint probability $\Pr(k_i, k_j)$.

- Prior works have primarily focused on the **combination process**, measuring how typical or atypical a resulting combination is.
 - e.g., atypicality, often interpreted as a proxy for novelty or interdisciplinarity, is shown to be associated with higher citation impact.
- They have largely overlooked the **retrieval process** that precedes it.
- Although some studies have operationalized **retrieval process**, they rarely account for individual researcher contexts.
 - i.e., they assume that knowledge components are equally accessible to all researchers, regardless of their prior trajectories and expertise.
- Why not retrieve any knowledge component to make an ideal combination?

Knowledge recombination is not only driven by the combination process, but also constrained by the retrieval process.

- **Retrieval process:** search/retrieve existing knowledge components.
 - e.g., **physicists** are more likely to retrieve **physics** than **biology**.
 - Conditional probability $\Pr(k_i | H_a)$.
- **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 $\Pr(k_i, k_j)$.
- How does the retrieval process constrain the combination process? When atypical combinations are built, what happens in the retrieval process?
 - e.g., who would conduct research combining **physics** and **biology**?
 - **Physicists**, **biologists**, or **interdisciplinary** researchers who are familiar with both domains?
 - $\Pr(k_i | H_a), \Pr(k_j | H_a) \longleftrightarrow \Pr(k_i, k_j)$

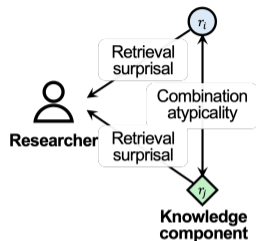
How does the retrieval process constrain the combination process?

- **RQ1:** Is atypical combination produced by relying on familiar knowledge, or by leaping into unfamiliar knowledge?
 - **Expectation:** Atypical combinations are more likely to be constructed from components that are less likely to be retrieved (i.e., difficult to acquire).
- **RQ2:** Why do some atypical combinations become high-impact, disruptive, or consolidating?
 - **Expectation:** Atypical combinations are more impactful when they are built with unfamiliar knowledge components.
- We answer these questions by linking two dimensions of recombination.

- We analyze knowledge recombination in **scientific publications**.
 - Publication and citation data from OpenAlex and SciSciNet-v2.
- **Knowledge components:** journals cited by each paper.
 - Each paper is represented as a set of journals it cites: $p = \{r_1, r_2, \dots, r_n\}$.
 - Each cited journal r represents a knowledge component.
- **Researcher trajectories:** author-level publication histories ordered over time.
 - Prior publication history is represented by the sequence of prior papers: $H_a = \{p_1, p_2, \dots, p_m\}$.
 - Only include first-authored publications to capture the researcher's own trajectory.
- **Sample restrictions:**
 - Papers with 20–50 references published between 2000–2024.
 - Papers must have enough references to compute combination atypicality, but ones with excessively long reference lists (such as review papers) excluded.

Two Dimensions of Recombination

- We link two dimensions of recombination: retrieval and combination.
- **Retrieval surprisal** (local accessibility): how unexpected a cited journal is given a researcher's prior publication history.
- **Combination atypicality**: how unusual a cited-journal pair is in the global citation network.



	Retrieval / accessibility	Combination / atypicality
Unit	Cited journal (component)	Cited-journal pair (combination)
Reference frame	Researcher's prior trajectory H_a	Population-level citation network
Probability	$\Pr(r_i H_a)$	$\Pr(r_i, r_j)$ relative to marginals
Question	Could this researcher reach it?	Is this pair uncommon in science?
Interpretation	Local accessibility / retrieval cost	Global uncommonness / rarity

Measurement: Trajectory-conditioned Local Accessibility

- **Retrieval surprisal** (local accessibility): how unexpected a cited journal is given a researcher's prior publication history.
- A sequential model predicts the journals a researcher is likely to cite in the next paper, conditional on their prior publication history, thus estimating $\Pr(r|H_a)$.
- For each observed cited journal r_i , **retrieval surprisal** is defined as:

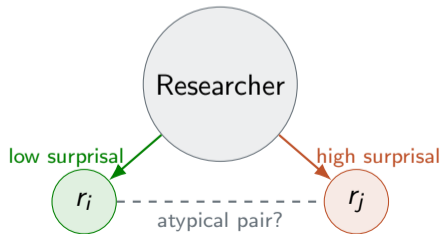
$$S(r|H_a) = -\log \Pr(r|H_a)$$

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

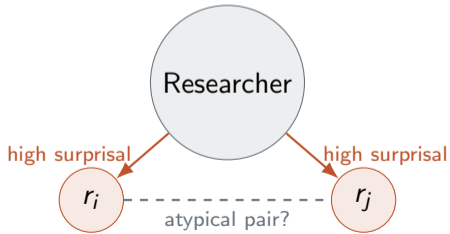
Pair typology: Local Accessibility

	Low S (accessible)	High S (distant)
Low S	Within-trjectory	Pivoting
High S	Pivoting (anchored exploration)	Leaping (unanchored exploration)

Pivoting (from an anchor)



Leaping



Measurement: Global Combination Atypicality

- **Combination atypicality**: how unusual a cited-journal *pair* is in the global citation network.
- For each cited-journal pair (r_i, r_j) , we calculate the **combination atypicality** using normalized pointwise mutual information (NPMI):

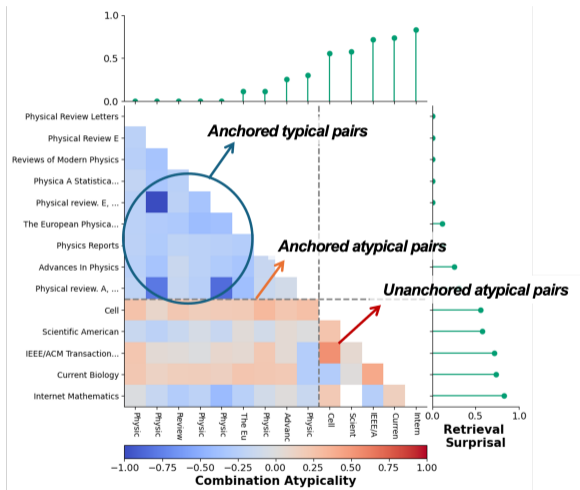
$$A(r_i, r_j) = -\text{npmi}(r_i, r_j) = \frac{\log(\Pr(r_i) \Pr(r_j)) - \log \Pr(r_i, r_j)}{-\log \Pr(r_i, r_j)}$$

- Low (negative) value means the pair is commonly observed together; high (positive) value means the pair is rarely observed together, relative to the expected co-occurrence based on their individual frequencies.

Pair typology: Local Accessibility & Global Atypicality

	Typical pair $A \leq 0$	Atypical pair $A > 0$
Within-trjectory (both accessible)	Anchored typical pair	Anchored atypical pair
Pivoting (asymmetric)	Anchored typical pair	Anchored atypical pair
Leaping (both distant)	Unanchored typical pair	Unanchored atypical pair

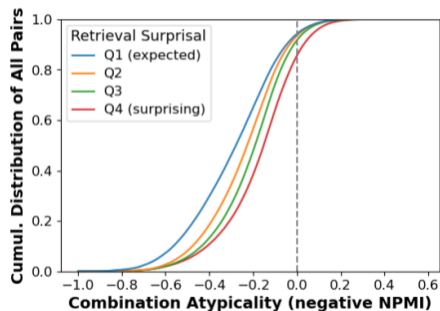
Illustration of Pair-Level Structure



Result 1: Pair-Level Structure

Research Question 1

Is atypical combination produced by relying on familiar knowledge, or by leaping into unfamiliar knowledge?



- Atypical pairs are more frequently observed among hard-to-retrieve components.

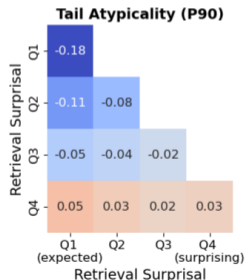
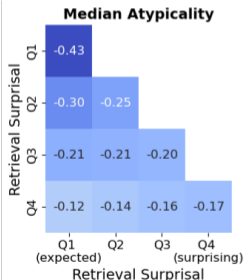
Result 1: Pair-Level Structure

Asymmetric retrieval profiles

Research Question 1

Is atypical combination produced by relying on familiar knowledge, or by leaping into unfamiliar knowledge?

Routine low-low	Pivot low-high
Pivot high-low	Leap high-high



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

Result 1: Pair-Level Structure

Regression analysis

$$A(r_i, r_j) = \beta_0 + \beta_1 S(r_i | H_a) + \beta_2 S(r_j | H_a) + \beta_3 S(r_i | H_a) S(r_j | H_a) + \epsilon_{ij,p}$$

Term	Estimate	Interpretation
$S(r_i H_a)$	0.294***	A distant component increases atypicality.
$S(r_j H_a)$	0.294***	-
$S(r_i H_a) \times S(r_j H_a)$	-0.357***	Both-high is less distinctive than low-high.

From Pair-Level Structure to Paper-Level Outcomes

Research Question 2

Why do some atypical combinations become high-impact, disruptive, or consolidating?

Dimension	Paper-level predictor	Definition
Retrieval profile	Exploration breadth Fraction of anchors	Mean retrieval surprisal across components. Share of components with low retrieval surprisal ($S \leq Q1$).
Global combination	Median pair atypicality Fraction of atypical pairs	Median atypicality across all component pairs. Share of component pairs with positive atypicality ($A > 0$).
Pair topology	Fraction of anchored typical pairs	Typical pairs ($A \leq 0$) containing at least one anchor.
	Fraction of anchored atypical pairs	Atypical pairs ($A > 0$) containing at least one anchor.
	Fraction of unanchored atypical pairs	Atypical pairs ($A > 0$) with both components distant.
	Fraction of unanchored typical pairs	Typical pairs ($A \leq 0$) with both components distant.

Paper-Level Regression Results

Predictor	(1) Impact	(2) Impact	(3) Impact	(4) Disrupt.	(5) Disrupt.
Mean retrieval surprisal	-0.458***	0.832***	0.825***	-0.001***	-0.001***
Fraction of anchors		1.287***		-0.002***	
Median atypicality	-1.229***	-2.558***	-2.572***	-0.000	-0.000
Fraction of atypical pairs		2.813***		0.001***	
Fraction of anchored atypical pairs			4.367***		-0.003***
Fraction of unanchored atypical pairs			2.608***		0.002***
Fraction of anchored typical pairs			1.249***		-0.002***
Pseudo R^2 / R^2	0.016	0.022	0.022	0.009	0.009

Logistic regressions predict top-5% citation impact (model 1–3).

OLS regressions predict disruption index (model 4–5).

Controls: number of reference pairs and team size. *** $p < 0.01$.

Result 2: Paper-Level Citation Impact

Exploration breadth and anchors

Predictor	Model 1	Model 2	Model 3
Exploration breadth	-0.458***	0.832***	0.825***
Fraction of anchors		1.287***	
Median atypicality	-1.229***	-2.558***	-2.572***
Fraction of atypical pairs		2.813***	
Fraction of anchored atypical pairs			4.367***
Fraction of unanchored atypical pairs			2.608***
Fraction of anchored typical pairs			1.249***

- Exploration breadth becomes positively associated with impact once the presence of anchors is accounted for.
- **The apparent cost of exploration (also known as “pivot penalty”) is largely associated with losing cognitive anchors**, rather than average retrieval difficulty of components.

Result 2: Paper-Level Citation Impact

Anchored atypical pairs

Research Question 2-1

Why do some atypical combinations become high-impact?

	Typical pair	Atypical pair
Unanchored pair	0 baseline	2.608*** unanchored atypical
Anchored pair	1.249*** anchored typical	4.367*** anchored atypical

- Fraction of anchored atypical pairs exceeds the additive effects of (fraction of) anchors and atypical pairs: $4.367 > 1.249 + 2.608$.
- **Atypical combinations appears to be more impactful when they are built from pivoting with familiar anchors, rather than leaping without any anchor.**

Result 3: Disruption

Research Question 2-2

Why do some atypical combinations become disruptive or consolidating?

Pair type	Top-5% citation impact	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)

- Influence of atypical combinations appears to differ by retrieval structure.
- **Anchors may make atypical combinations easier to integrate into existing knowledge streams, promoting consolidation.**
- **Unanchored atypical combinations may have a difficult role in disrupting existing streams.**

Claim

Retrieval structure of knowledge components matters for the combination process and its consequences.

- **Pair-level structure:** Atypical pairs are more likely to be observed in pivoting with anchors, rather than leaping.
- **Citation impact:** Anchored atypical pairs have a stronger association with high impact than unanchored atypical pairs.
- **Disruption:** Anchored atypicality pairs consolidates; unanchored atypicality disrupts.

Current scope

Retrieval surprisal is measured from first-author trajectories; knowledge components are represented as cited journals.

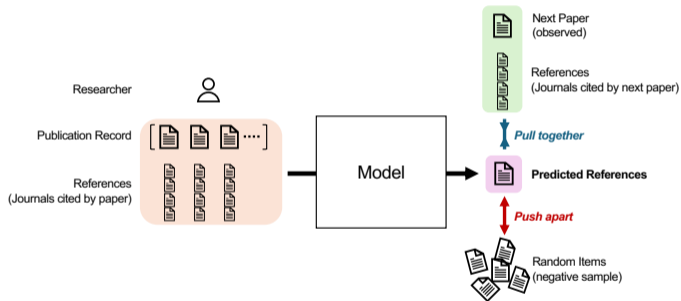
- **Team-level retrieval:** Scientific recombination is collaborative; exploring team-level retrieval profiles and their interplay with combination atypicality is an important next step.
- **Diverse knowledge units:** Future robustness checks will use alternative representations such as topics, concepts, or semantic embeddings.
- **Other domains:** The framework can be extended to other domains such as patents, to examine whether anchored recombination is a generalizable phenomenon in innovation processes.

Thank you for your attention!
Questions and feedback are very welcome.

saeyoung.choi@snu.ac.kr

Backup: Sequential model details

a



b

