Tracing Recombinational Creativity: A Multi-Level Framework for Measuring Novelty in Creative Processes

YULIN YU, Northwestern University, University of Michigan, USA

Creativity is often understood as the recombination of existing ideas into novel configurations. While prior research has shown that atypical combinations can drive breakthrough innovation, most quantitative approaches measure recombination at the level of finished artifacts—papers, patents, or products—treating novelty as an outcome rather than a process. As a result, we lack a mechanistic understanding of how recombination unfolds and why certain combinational strategies succeed while others fail. The emergence of creativity support tools, particularly AI-based systems, offers a new opportunity: rich activity traces now allow us to observe creative recombination as it unfolds through prompts, revisions, branching edits, and component reuse. In this position paper, I argue that recombinational novelty should be reconceptualized as a traceable, multi-level process. I illustrate this perspective with three complementary measurement approaches using natural language processing and network analysis: surface-level recombination via TF–IDF techniques, conceptual recombination via embedding methods, and structural recombination via the Rao–Stirling diversity framework. Together, these approaches demonstrate how recombination can be quantified across representational levels and linked to both creative risk and impact.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; • **Data mining**;

ACM Reference Format:
Yulin Yu. 2018. Tracing Recombinational Creativity: A Multi-Level Framework for Measuring Novelty in Creative Processes. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Creativity is often described as the recombination of existing ideas into new configurations. Prior work in innovation studies shows that unusual combinations of knowledge elements are frequently associated with high-impact discoveries [2, 3, 5]. Yet recombination is inherently risky: while distant combinations can produce breakthrough outcomes, most attempts fail. This tension raises a central question: which kinds of recombination lead to creative success, and how do they unfold in practice?

Most quantitative studies examine recombination at the level of finished artifacts—such as papers, patents, or products—by measuring whether their components form atypical combinations. While these approaches reveal when outputs are unusual, they provide limited insight into how novelty is constructed. In particular, they treat recombination as a property of outcomes rather than as a process through which creative elements are assembled and refined.

The emergence of creativity support tools, especially AI-based systems, opens new opportunities for studying recombination more directly. These tools generate detailed traces of creative activity—including prompts, edits, branching

Author's Contact Information: Yulin Yu, yulinyu@umich.edu, Northwestern University, University of Michigan, Ann Arbor, Michigan, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

explorations, and component reuse—that capture the sequence of decisions underlying creative work. Such activity traces make it possible to observe recombination as it unfolds rather than inferring it solely from final outputs.

In this position paper, I argue that recombinational novelty can be reconceptualized as a traceable, multi-level process. Drawing on examples from my prior work, I illustrate how recombinational creativity can be quantified across different domains by modeling the entities being combined and the structures through which they interact. This perspective moves toward a more mechanistic understanding of creativity and provides a set of quantitative tools for studying creative processes in AI-mediated environments.

2 Conceptual Framework

Under the recombinant view of innovation, creativity emerges when existing elements are combined in configurations that differ from prior patterns of use [2, 6]. Operationalizing this idea requires identifying the relevant "building blocks" in a domain and measuring how distinct their combinations are relative to historical baselines.

Importantly, recombination can occur at multiple representational levels. In sequence-based cultural artifacts such as songs, creative elements may consist of lyrics or chords. In scientific research, recombination may occur through the reuse of datasets across different conceptual contexts. More broadly, novelty can arise from structurally unusual combinations of resources within a work.

Guided by this perspective, we present three complementary approaches for quantifying recombinational creativity. First, we measure distinctiveness in observable components using TF–IDF-style representations, capturing how unusual a configuration of elements is relative to contemporary artifacts. Second, we measure recombination in conceptual space using semantic embedding models, which quantify the extent to which existing resources—such as datasets—are reused in semantically distant contexts. Third, we measure structural novelty using a diversity framework based on the Rao–Stirling index, which captures how atypical a set of components is relative to historical co-usage patterns.

Together, these approaches operationalize recombinational creativity across three levels: surface composition, conceptual recontextualization, and structural combination. By applying these measures, we demonstrate how theories of recombinant innovation can be translated into scalable quantitative indicators across diverse creative domains.

3 Illustrative Examples

To demonstrate how the conceptual framework can be operationalized in practice, we present three illustrative examples drawn from prior work. Each example corresponds to one of the three levels of recombinational creativity described in the conceptual framework: (1) surface composition, (2) conceptual recontextualization, and (3) structural combination. Together, these cases illustrate how the same theoretical idea—novel recombination of existing elements—can be translated into quantitative measures across different domains and representations.

3.1 Surface-Level Recombination in Sequence-Based Artifacts

Our first example illustrates recombinational creativity at the *surface composition* level. In this setting, the recombinable units are observable elements within a creative artifact. We study sequence-based artifacts in music—specifically song lyrics and chords—where creative outputs can be decomposed into recombinable components [9]. Rather than treating a song as a monolithic artifact, we represent it as a configuration of linguistic and harmonic elements and measure how distinct this configuration is relative to contemporaneous works. We quantify *Lyrics Uniqueness* as word-pair or phrase-pair recombination using a TF–IDF vector space representation. Each song is treated as a document, and its lyrics are represented as a weighted term vector after standard preprocessing (removing stopwords, punctuation, and non-English terms). TF–IDF downweights commonly used words and highlights distinctive vocabulary, allowing us to capture how a song recombines linguistic elements relative to other songs released in the same calendar year. For each song , we compute the cosine similarity between its TF–IDF vector and the average TF–IDF vector of all other songs released in the same year. Regarding*Chord Uniqueness*, We apply an analogous approach to harmonic structure. Instead of modeling full chord sequences, we represent each song as a TF–IDF vector of chord frequencies (e.g., "C Major," "E Minor"), treating chords as recombinable units. Specific details of the measurement, equations, and validation can be found in [9]. Together, these measures capture recombination at the level of observable components, operationalizing the surface composition layer of the conceptual framework.

3.2 Conceptual Recombination

Our second example illustrates recombination at the level of *conceptual recontextualization*. Here the building blocks are not surface lexical elements but prior knowledge and ideas used in scientific research.

We operationalize recombinational creativity through *conceptual repurposing*: the extent to which a focal paper reuses existing knowledge in a semantically novel research context. For each focal paper, we identify prior papers published at least one year earlier that represent its intellectual predecessors (e.g., through citation links). We then construct Sentence-BERT embeddings based on paper titles to represent studies in a high-dimensional semantic space. Recombinational distance is measured as the semantic distance between the focal paper and its cited predecessors.

Intuitively, if prior knowledge is reused in conceptually similar contexts, the semantic distance will be small, indicating incremental reuse. Larger distances suggest that existing ideas are being repurposed across different theoretical framings, research questions, or disciplinary domains.

This embedding-based approach captures recombination at a deeper representational level than surface lexical variation, revealing how existing knowledge is repurposed into new intellectual contexts[8]..

Our second example illustrates recombination at the level of *conceptual recontextualization*. Here the building blocks are not surface elements but knowledge resources—in this case, concept used in scientific research.

We operationalize recombinational creativity through *repurposing*: the extent to which a focal paper applies an existing dataset in a semantically novel research context. For each focal paper, we identify prior papers published at least one year earlier that used overlapping datasets. We then construct Sentence-BERT embeddings based on paper titles to represent studies in a high-dimensional semantic space. Recombinational distance is measured as the semantic distance between the focal paper and its dataset-sharing predecessors [8].

Intuitively, if a dataset is reused in conceptually similar contexts, the semantic distance will be small, indicating incremental reuse. Larger distances suggest that the same dataset is being recombined with different theoretical framings, research questions, or disciplinary domains. We normalize this measure to range from 0 to 1, where higher values indicate greater conceptual recombination.This embedding-based approach captures recombination at a deeper representational level than surface lexical variation, revealing how existing data resources are integrated into new intellectual contexts.

3.3 Structural Recombination

Our third example illustrates recombination at the level of *structural combination*. Rather than measuring conceptual distance across papers, we examine how datasets are combined within a single paper. We quantify this structural novelty using a general diversity framework based on the Rao–Stirling index [1, 4, 7]. Combinations of datasets that rarely appear together in prior research are considered more atypical and therefore more recombinationally novel.

We identify the set of elements used in each paper and examine how these elements are combined within a single study. To quantify how atypical these combinations are, we construct a co-usage matrix across the corpus that records how frequently pairs of elements appear together in prior work. For each element, we build a usage vector indicating the papers in which it has previously appeared. The similarity between two elements is then computed as the cosine similarity between their usage vectors [10].

Following the Rao–Stirling diversity framework, we quantify recombinational atypicality based on how rarely the elements used in a focal paper have been combined in prior research. Papers that combine elements with historically low co-usage similarity are considered to exhibit greater structural recombination. In empirical analyses, this measure is normalized to facilitate comparison across papers. Specific details of the measurement, equations, and validation can be found in [10].

4 Conclusion and Open Questions

Recombinational novelty has long been recognized as a central mechanism of creativity, yet empirical studies have largely measured it at the level of completed artifacts. In this position paper, I argue that recombination can instead be understood as a traceable, multi-level process. By operationalizing recombination across component-level distinctiveness (TF–IDF), semantic recontextualization (embedding-based distance), and structural pairing atypicality (Rao–Stirling diversity), we move toward a more mechanistic account of how novelty is constructed. These approaches illustrate how creative building blocks—words, chords, datasets, or other modular elements—can be analyzed to reveal distinct combinational strategies and their associated risk–reward profiles.

At the same time, recombination represents only one pathway to creativity. Some forms of creative change arise not from combining existing elements but from introducing ideas that disrupt established trajectories or render prior approaches obsolete. Such disruptive creativity may involve conceptual breaks, paradigm shifts, or entirely new problem framings that cannot be fully captured through recombinational metrics alone. Recognizing this distinction suggests that creativity should not be treated as a single dimension but rather as a multidimensional phenomenon encompassing recombination, disruption, exploration, and other forms of novelty.

The growing availability of creative activity traces—particularly in AI-mediated creativity support tools—opens new opportunities for studying these processes. Instead of inferring novelty from static outputs, we can observe how creative ideas emerge through sequences of edits, branching explorations, and human–AI interactions. This shift raises several important questions: How can recombinational strategies be detected from trace data in real time? How do different creative mechanisms—such as recombination and disruption—interact within the same creative process? How does AI influence the balance between exploratory search and integrative synthesis? And how can we design measurement frameworks that capture multiple dimensions of creativity simultaneously?

Addressing these questions may help bridge innovation theory and HCI by moving beyond artifact-level indicators toward dynamic measurements of creative processes. Ultimately, developing multidimensional and trace-based metrics of creativity will be essential for understanding—and designing—the next generation of human–AI creative systems.

References

[1] Loredana Cassi, Antoine Morrison, and Fabien Ter Wal. 2017. The Evolution of the Interdisciplinary Nature of Knowledge Production: Evidence from Three Fields. *Research Policy* 46, 4 (2017), 747–762. doi:10.1016/j.respol.2017.01.009

[2] Lee Fleming. 2001. Recombinant Uncertainty in Technological Search. *Management Science* 47, 1 (2001), 117–132. doi:10.1287/mnsc.47.1.117.10671

[3] Joseph A. Schumpeter. 1934. *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.

[4] Andy Stirling. 2007. A General Framework for Analysing Diversity in Science, Technology and Society. *Journal of The Royal Society Interface* 4, 15 (2007), 707–719. doi:10.1098/rsif.2007.0213

[5] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical Combinations and Scientific Impact. *Science* 342, 6157 (2013), 468–472. doi:10.1126/science.1240474

[6] Martin L. Weitzman. 1998. Recombinant Growth. *Quarterly Journal of Economics* 113, 2 (1998), 331–360. doi:10.1162/003355398555595

[7] Yang Yang, Dashun Wang, and Brian Uzzi. 2022. Quantifying the Dynamics of Interdisciplinary Research. *Nature Communications* 13 (2022), 1234. doi:10.1038/s41467-022-XXXXX

[8] Yulin Yu, Yong-Yeol Ahn, and Daniel M. Romero. [n. d.]. Dataset Re-Purposing Fuels AI Breakthroughs. ([n. d.]). Working paper (unpublished manuscript).

[9] Yulin Yu, Pui Yin Cheung, Yong-Yeol Ahn, and Paramveer S Dhillon. 2023. Unique in what sense? Heterogeneous relationships between multiple types of uniqueness and popularity in music. In *Proceedings of the international AAAI conference on web and social media*, Vol. 17. 914–925.

[10] Yulin Yu and Daniel M Romero. 2024. Does the use of unusual combinations of datasets contribute to greater scientific impact? *Proceedings of the National Academy of Sciences* 121, 41 (2024), e2402802121.