

Atypical Combinations & Scientific Impact

Uzzi et al. 2013

Monday, June 6, 2022

Objectives

- Where is scientific innovation found?
- Part I: Novelty v Conventionality
- Part II: Highly Cited Works
- Part III: Collaboration

Methods & Data

Overview

- 17.9 million Web of Science Articles
- Count frequency of co-citation pairs
- Compared against random combinations
- Counts of paper pairs → Journal Pairs
- 15,613 journals in WoS

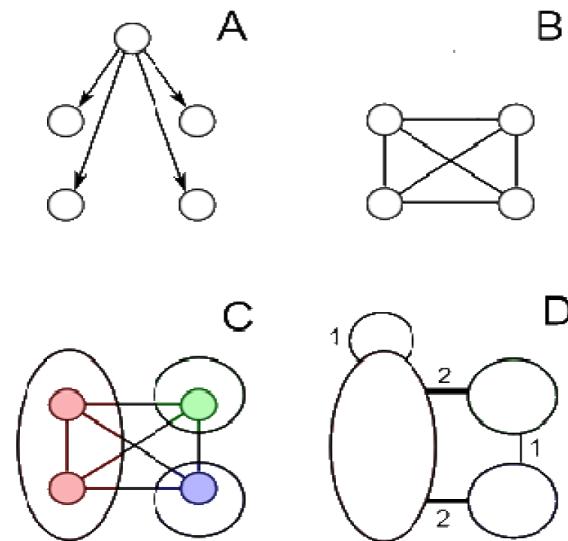
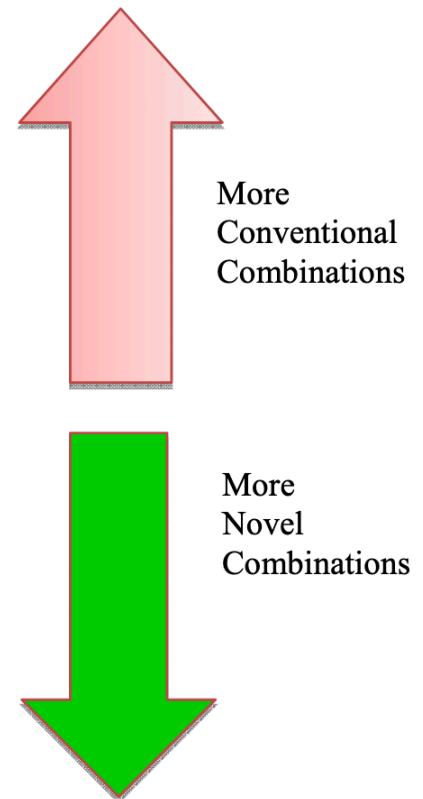


Table S1. Examples of Journal Pair Frequencies for Illustrative Paper

Journal Pairs	Observed	Expected	Z-score
Tetrahedron - Tetrahedron	5071	151.89	637.77
Experientia - Experientia	1159	109.59	95.07
Tetrahedron - Experientia	454	256.06	21.55
Experientia - Tetrahedron	661	481.07	6.88
Lett			
Z-score of Zero means obs is as likely as chance			0.0
Chem Phar Bull - Life Sci	114	151.19	-2.4
Life Sci - R J Royal Neth C	16	45.45	-4.82
Life Sci – Tetrahedron	36	315.78	-17.67
Life Sci – J Organic Chemistry	166	813.72	-24.21
J Am Chem Soc - Life Sci	469	3147.65	-45.07



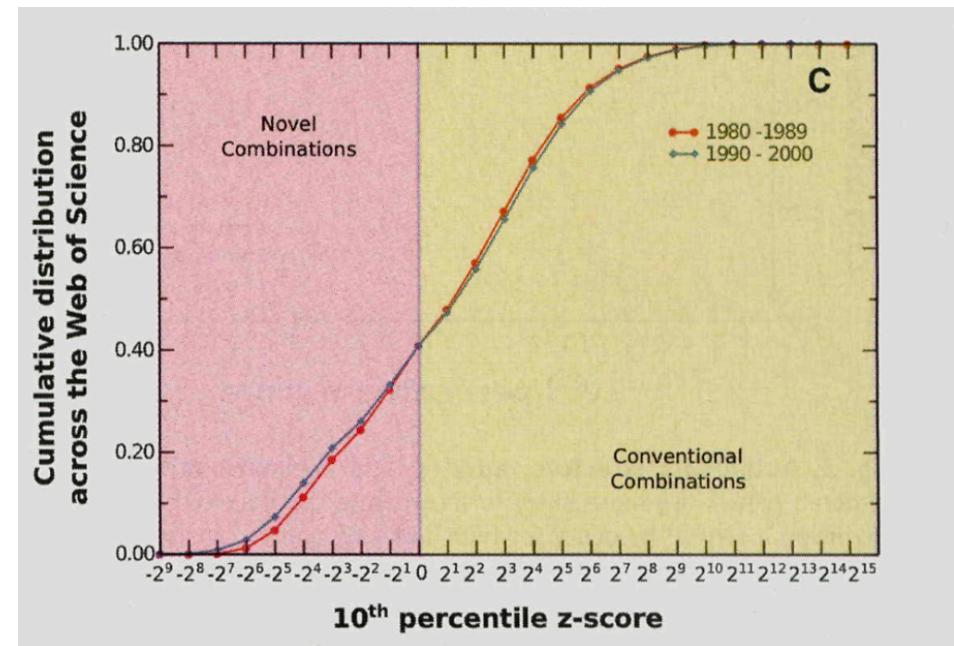
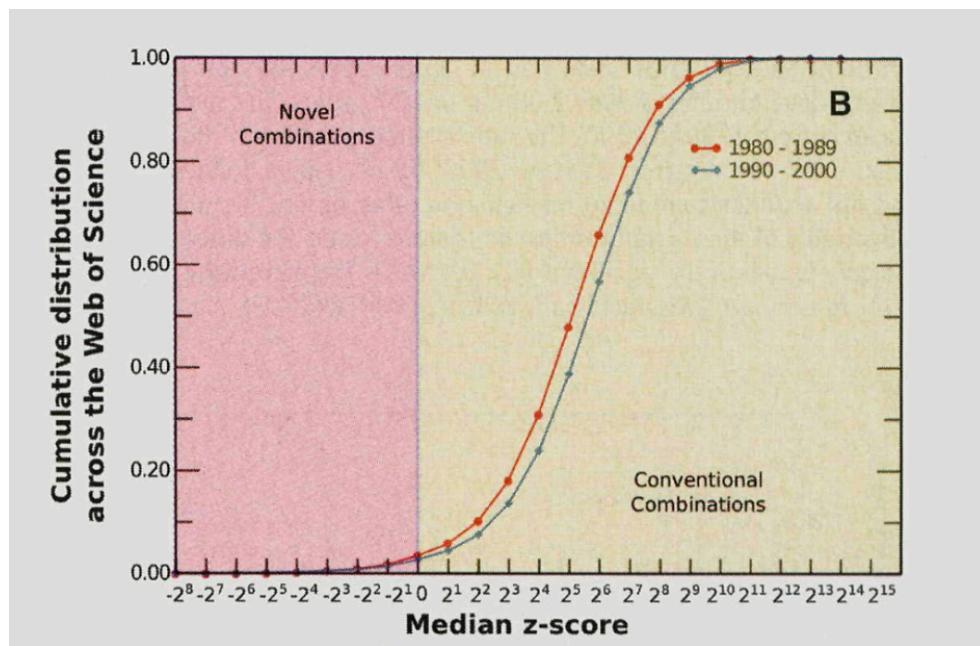
Methods & Data

Z-Scores

- Each pair gets a z-score
 - Positive = conventional pair
 - Negative = novel pair
- Summarizing a paper as novel or conventional
 - Median z-score
 - 10th percentile z-score (“left tail”)

Results

Z-scores



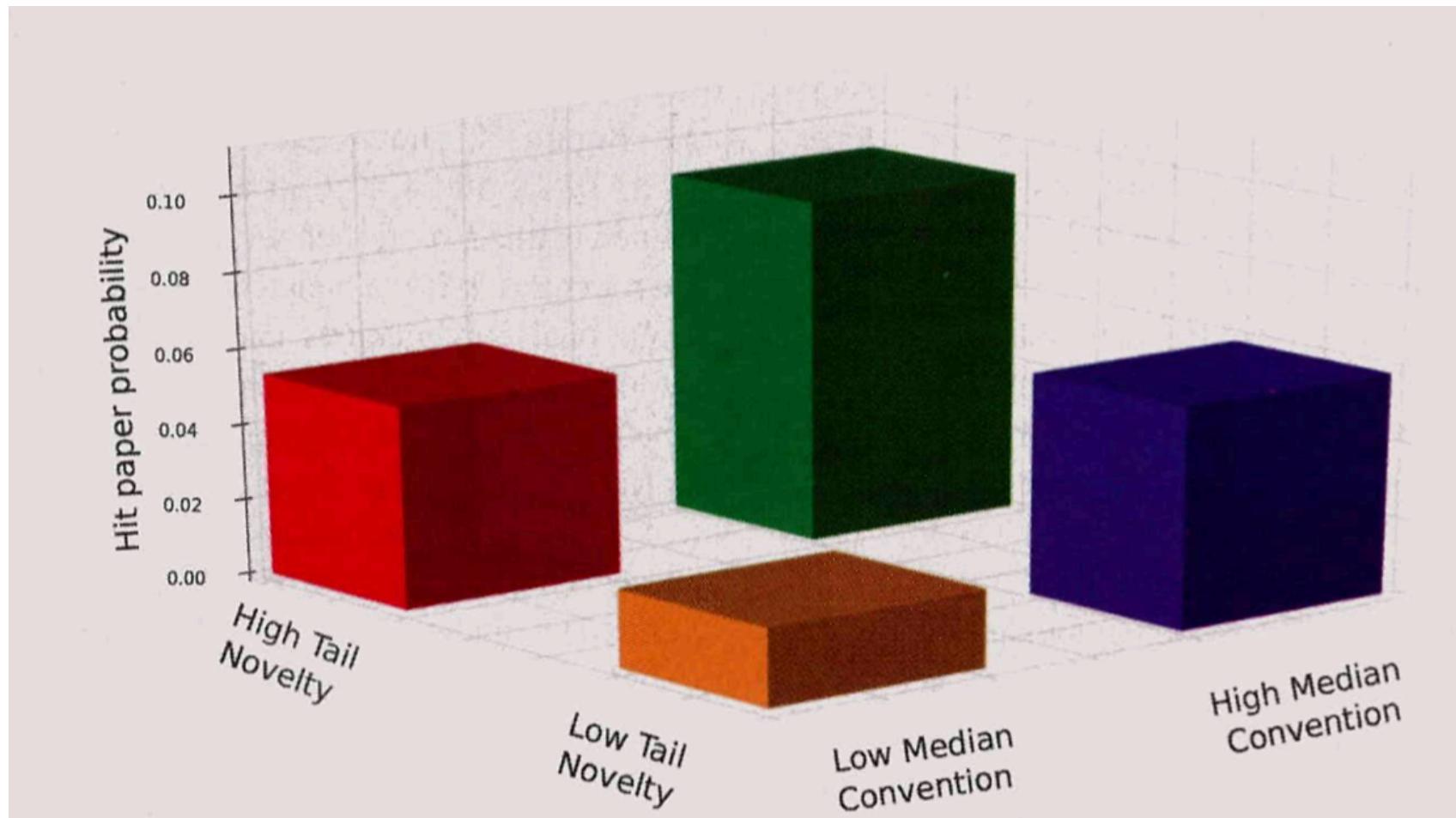
Results

- Overall: Papers prefer conventionality
- Even the most novel combinations fall on more conventional side
- Can we predict which combinations will lead to a “hit paper?”

Part II

Predictability

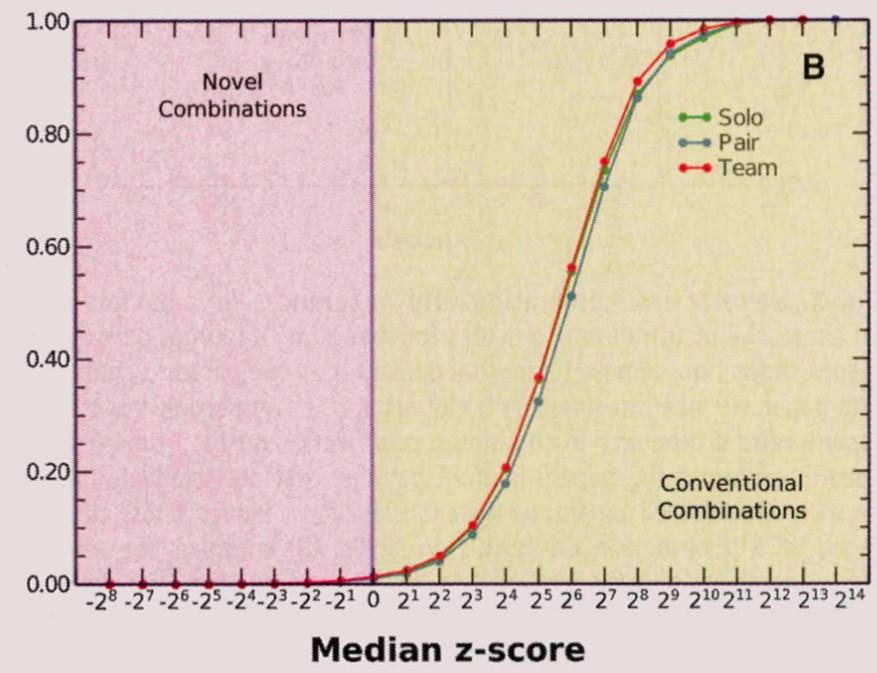
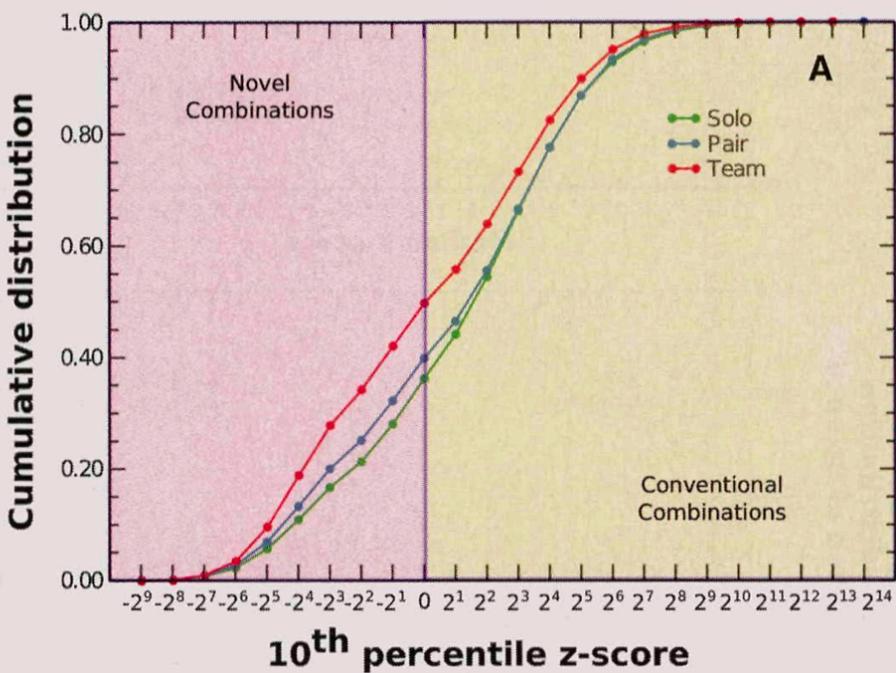
- “Median conventionality” = is the median z-score in the top half or bottom half
 - High conventionality can mean anywhere from just above average to most conventional
 - Low conventionality (relative) can still indicate conventionality score well above 0
- “Tail Novelty” = does the z-score for the most novel 10% of papers run positive or negative?



Part III

Teams

- Collaboration and novelty
- Single author, dual author, and team authored papers



Conclusions

(up for critique)

- More likely to be highly cited:
 - Team authored papers
 - High tail novelty and high median convention (max 85-95th percentile)

Questions

- Were the original questions answered?
- Are journal pairs a good way to measure novelty?
- Low/High median convention
 - Should this value have been split into halves?
- Authors claim these methods can be applied to any “disciplines, papers, or topics within papers.”
 - Is this true?