Atypical Combinations and Scientific Impact

Brian Uzzi *et al.*

range of likely core temperatures. However, further investigations into multicomponent systems are essential to fully understand their effect on the elastic properties of the core. Overall, our results demonstrate that the inner core is likely to be in the strongly nonlinear regime; hence, there is no need to invoke special circumstances such as strong anelasticity, partial melts, or combinations of crystalline phases in order to match the observed seismic velocities and densities of the inner core.

References and Notes

1. F. Birch, *J. Geophys. Res.* **69**, 4377–4388 (1964).
2. J.-P. Poirier, *Phys. Earth Planet. Inter.* **85**, 319–337 (1994).
3. A. Cao, B. Romanowicz, N. Takeuchi, *Science* **308**, 1453–1455 (2005).
4. L. Vočadlo, *Earth Planet. Sci. Lett.* **254**, 227–232 (2007).
5. X. Sha, R. E. Cohen, *Phys. Rev. B* **81**, 094105–094110 (2010).
6. X. Sha, R. E. Cohen, *Geophys. Res. Lett.* **37**, L10302–L10305 (2010).
7. D. Antonangeli *et al.*, *Earth Planet. Sci. Lett.* **225**, 243–251 (2004).
8. A. P. Kantor *et al.*, *Phys. Earth Planet. Inter.* **164**, 83–89 (2007).
9. B. Martorell, J. Brodholt, I. G. Wood, L. Vočadlo, *Earth Planet. Sci. Lett.* **365**, 143–151 (2013).
10. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
11. D. Antonangeli *et al.*, *Earth Planet. Sci. Lett.* **295**, 292–296 (2010).
12. Z. Mao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10239–10244 (2012).
13. M.-H. Nadal, P. Le Poac, *J. Appl. Phys.* **93**, 2472–2480 (2003).
14. M.-H. Nadal, C. Hubert, G. Ravel-Chapuis, *J. Alloys Compds.* **444–445**, 265–267 (2007).
15. D. Alfè, G. D. Price, M. J. Gillan, *Phys. Rev. B* **65**, 165118 (2002).
16. D. Alfè, *Phys. Rev. B* **79**, 060101–060104 (2009).
17. E. Sola, D. Alfè, *Phys. Rev. Lett.* **103**, 078501–078504 (2009).
18. G. Morard, J. Bouchet, D. Valencia, S. Mazevet, F. Guyot, *High Energy Density Phys.* **7**, 141–144 (2011).
19. See supplementary materials on *Science* Online.
20. L. Vočadlo, D. Dobson, I. G. Wood, *Earth Planet. Sci. Lett.* **288**, 534–538 (2009).
21. Y. P. Varshni, *Phys. Rev. B* **2**, 3952–3958 (1970).
22. M. W. Guinan, D. J. Steinberg, *J. Phys. Chem. Solids* **35**, 1501–1512 (1974).
23. D. R. Nelson, *Defects and Geometry in Condensed Matter Physics* (Cambridge Univ. Press, Cambridge, 2002).
24. V. Sorkin, E. Polturak, J. Adler, *Phys. Rev. B* **68**, 174102–174107 (2003).
25. V. Sorkin, E. Polturak, J. Adler, *Phys. Rev. B* **68**, 174103–174109 (2003).
26. K. Lu, Y. Li, *Phys. Rev. Lett.* **80**, 4474–4477 (1998).
27. F. Delogu, *J. Phys. Chem. B* **110**, 12645–12652 (2006).
28. F. Delogu, *J. Phys. Chem. B* **110**, 3281–3287 (2006).
29. F. Delogu, *J. Phys. Condens. Matter* **18**, 5639–5653 (2006).
30. G. Manai, F. Delogu, *Physica B* **392**, 288–297 (2007).
31. G. Manai, F. Delogu, *J. Mater. Sci.* **42**, 6672–6683 (2007).
32. J. P. Poirier, *Introduction to the Physics of the Earth's Interior* (Cambridge Univ. Press, Cambridge, 2000), pp. 230–244.

Atypical Combinations and Scientific Impact

Brian Uzzi,[1,2] Satyam Mukherjee,[1,2] Michael Stringer,[2,3] Ben Jones[1,4]*

Novelty is an essential feature of creative ideas, yet the building blocks of new ideas are often embodied in existing knowledge. From this perspective, balancing atypical knowledge with conventional knowledge may be critical to the link between innovativeness and impact. Our analysis of 17.9 million papers spanning all scientific fields suggests that science follows a nearly universal pattern: The highest-impact science is primarily grounded in exceptionally conventional combinations of prior work yet simultaneously features an intrusion of unusual combinations. Papers of this type were twice as likely to be highly cited works. Novel combinations of prior work are rare, yet teams are 37.7% more likely than solo authors to insert novel combinations into familiar knowledge domains.

Scientific enterprises are increasingly concerned that research within narrow boundaries is unlikely to be the source of the most fruitful ideas (*1*). Models of creativity emphasize that innovation is spurred through original combinations that spark new insights (*2–10*). Current interest in team science and how scientists search for ideas is premised in part on the idea that teams can span scientific specialties, effectively combining knowledge that prompts scientific breakthroughs (*11–15*).

Yet the production and consumption of boundary-spanning ideas can also raise well-known challenges (*16–21*). If, as Einstein believed (*21*), individual scientists inevitably become narrower in their expertise as the body of scientific knowledge expands, then reaching effectively across boundaries may be increasingly challenging (*4*), especially given the difficulty of searching unfamiliar domains (*17, 18*). Moreover, novel ideas can be difficult to absorb (*19*) and communicate, leading scientists to intentionally display conventionality. In his *Principia*, Newton presented his laws of gravitation using accepted geometry rather than his newly developed calculus, despite the latter's importance in developing his insights (*22*). Similarly, Darwin devoted the first part of the *Origin of Species* to conventional, well-accepted knowledge about the selective breeding of dogs, cattle, and birds. From this viewpoint, the balance between extending science with atypical combinations of knowledge while maintaining the advantages of conventional domain-level thinking is critical to the link between innovativeness and impact. However, little is known about the composition of this balance or how scientists can achieve it.

In this study, we examined 17.9 million research articles in the Web of Science (WOS) to see how prior work is combined. We present facts that indicate (i) the extent to which scientific papers reference novel versus conventional combinations of prior work, (ii) the relative impact of papers based on the combinations they draw upon, and (iii) how (i) and (ii) are associated with collaboration.

We considered pairwise combinations of references in the bibliography of each paper (*23, 24*). We counted the frequency of each co-citation pair across all papers published that year in the WOS and compared these observed frequencies to those expected by chance, using randomized citation networks. In the randomized citation networks, all citation links between all papers in the WOS were switched by means of a Monte Carlo algorithm. The switching algorithm preserves the total citation counts to and from each paper and the distribution of these citation counts forward and backward in time to ensure that a paper (or journal) with n citations in the observed network will have n citations in the randomized network. For both the observed and the randomized paper-to-paper citation networks, we aggregated counts of paper pairs into their respective journal pairs to focus on domain-level combinations (*24–26*). In the data, there were over 122 million potential journal pairs created by the 15,613 journals indexed in the WOS.

[1]Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. [2]Northwestern Institute on Complex Systems, Northwestern University, 600 Foster, Evanston, IL 60208, USA. [3]Datascope Analytics, 180 West Adams Street, Chicago, IL 60603, USA. [4]National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, USA.

*Corresponding author. E-mail: bjones@kellogg.northwestern.edu

Comparing the observed frequency with the frequency distribution created with the randomized citation networks, we generated a z score for each journal pair. This normalized measure describes whether any given pair appeared novel or conventional. Z scores above zero indicate pairs that appeared more often in the observed data than expected by chance, indicating relatively common or "conventional" pairings. Z scores below zero indicate pairs that appear less often in the observed WOS than expected by chance, indicating relatively atypical or "novel" pairings. For example, in the year 1980, the pairing *Tetrahedron* and *Experientia* had a high z score (21.55) indicating a conventional pairing, whereas *Tetrahedron* paired with *Life Sciences* had a negative z score (–17.67), indicating a pairing more unusual than chance. The supplementary materials detail these computations, the null model, and an illustrative example (table S1 and figs. S1 to S3).

As a simple validation of the z score measure, we found that journal pairs from the same WOS disciplinary designation had significantly higher z scores than did interdisciplinary journal pairs (table S3 and fig. S11). At the same time, only a minority (40.1%) of interdisciplinary journal pairs were novel, having z scores below zero

in the 1990s. This pattern indicates that observed journal pairings from the same WOS disciplines tend to be conventional, and interdisciplinary WOS journal pairings are less substantially conventional but still not consistently novel.
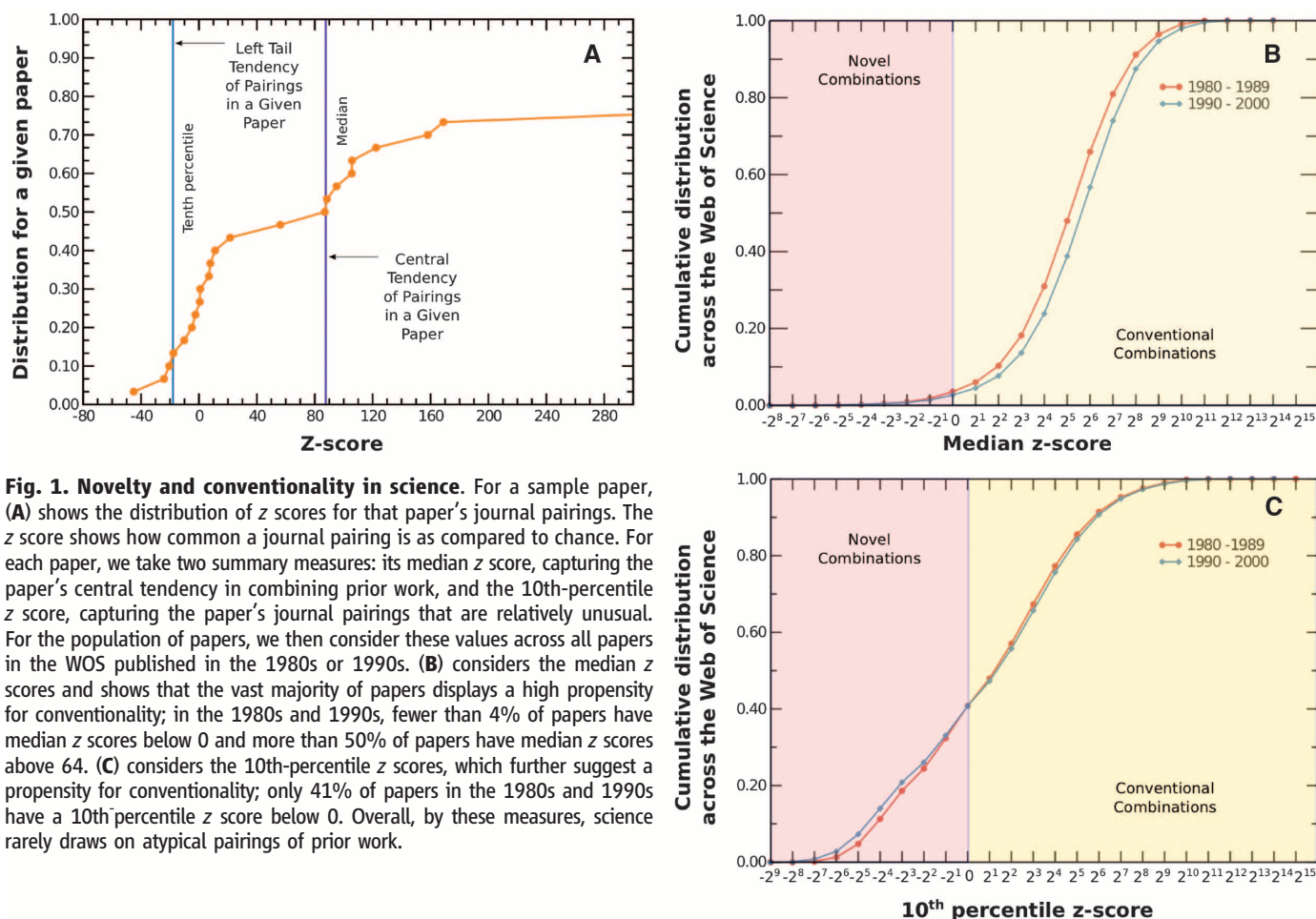
The above method assigns each paper a distribution of journal pair z scores based on the paper's reference list (Fig. 1A). To characterize a paper's tendency to draw together conventional and novel combinations of prior work, we examined two summary statistics. First, to characterize the central tendency of a paper's combinations, we considered the paper's median z score. The median allows us to characterize conventionality in the paper's main mass of combinations. Second, we considered the paper's 10th-percentile z score. The left tail allows us to characterize the paper's more unusual combinations, where novelty may reside.

We found that papers typically relied on very high degrees of conventionality. Figure 1B presents the distribution of papers' median z scores for the WOS in the indicated decades. Considering that a z score below zero represents a journal pair that occurs less often than expected by chance, the analysis of median z scores suggests very high degrees of conventionality. Half of the papers have median z scores exceeding

69.0 in the 1980s and 99.5 in the 1990s. Moreover, papers with a median z score below zero are rare. In the 1980s, only 3.54% of papers had this feature, whereas in the 1990s the percentage fell to 2.67%, indicating a persistent and prominent tendency for high conventionality.

Focusing on each paper's left tail combinations, we found that even among the paper's relatively unusual journal combinations, the majority of papers did not feature atypical journal pairs. Figure 1C shows that 40.8% of the papers in 1980s and 40.7% in the 1990s have a 10th-percentile z score below zero. Overall, by these measures, science typically relies on highly conventional combinations and rarely incorporates journal pairs that are uncommon compared to chance. Analyses in the supplementary materials (fig. S6) show that these empirical regularities for the WOS taken as a whole are largely replicated on a field-by-field basis and across time.

Our next finding indicates a powerful relationship between combinations of prior work and ensuing impact. Figure 2 presents the probability of a "hit" paper, conditional on the combination of its referenced journal pairs. Hit papers are operationalized as those in the upper 5th percentile of citations received across the whole data set, as measured by total citations through 8 years

Fig. 1. Novelty and conventionality in science. For a sample paper, (**A**) shows the distribution of z scores for that paper's journal pairings. The z score shows how common a journal pairing is as compared to chance. For each paper, we take two summary measures: its median z score, capturing the paper's central tendency in combining prior work, and the 10th-percentile z score, capturing the paper's journal pairings that are relatively unusual. For the population of papers, we then consider these values across all papers in the WOS published in the 1980s or 1990s. (**B**) considers the median z scores and shows that the vast majority of papers displays a high propensity for conventionality; in the 1980s and 1990s, fewer than 4% of papers have median z scores below 0 and more than 50% of papers have median z scores above 64. (**C**) considers the 10th-percentile z scores, which further suggest a propensity for conventionality; only 41% of papers in the 1980s and 1990s have a 10th-percentile z score below 0. Overall, by these measures, science rarely draws on atypical pairings of prior work.

after publication (the supplementary materials consider alternative definitions of hit papers). The vertical axis shows the probability of a hit paper conditional on a 2 × 2 categorization indicating the paper's (i) "median conventionality" (an indicator for whether the paper's median z score is in the upper or lower half of all median z scores) and (ii) "tail novelty" (an indicator for whether the paper's 10th-percentile z score is above or below zero).

Papers with "high median conventionality" and "high tail novelty" display a hit rate of 9.11 out of 100 papers, or nearly twice the background rate of 5 out of 100 papers. All other categories show significantly lower hit rates. Papers featuring high median conventionality but low tail novelty displayed hit rates of 5.82 out of 100 papers, whereas those featuring low median conventionality but high tail novelty display hit rates of 5.33 out of 100 papers. Finally, papers low on both dimensions have hit rates of just 2.05 out of 100.

Further analyses suggest a universality of these relationships for scientific work across time and fields. We considered the same relationships for different time periods (fig. S4), for different definitions of high-impact papers (fig. S5), and for each of 243 fields of science (fig. S6 and table S2). These analyses confirmed the findings above. Thus, novelty and conventionality are not opposing factors in the production of science; rather, papers with an injection of novelty into an otherwise exceptionally familiar mass of prior work are unusually likely to have high impact.

Collaboration is often claimed to produce more novel combinations of ideas (10–14), but the extent to which teams incorporate novel combinations across the universe of fields is unknown. Team-authored papers were more likely to show atypical combinations than were single- or pair-authored papers. Figure 3A shows that the distribution of 10th-percentile z scores shifted significantly leftward as the number of authors increased [Kolmogorov-Smirnov (K-S) tests indicate solo versus pair $P = 0.016$, pair versus team $P = 0.001$, team versus solo $P < 0.001$]. Papers written by one, two, or three or more authors showed high tail novelty in 36.1, 39.8, and 49.7% of cases, respectively, indicating that papers with three or more authors showed an increased frequency of high tail novelty over the solo-author rate by 37.7%.
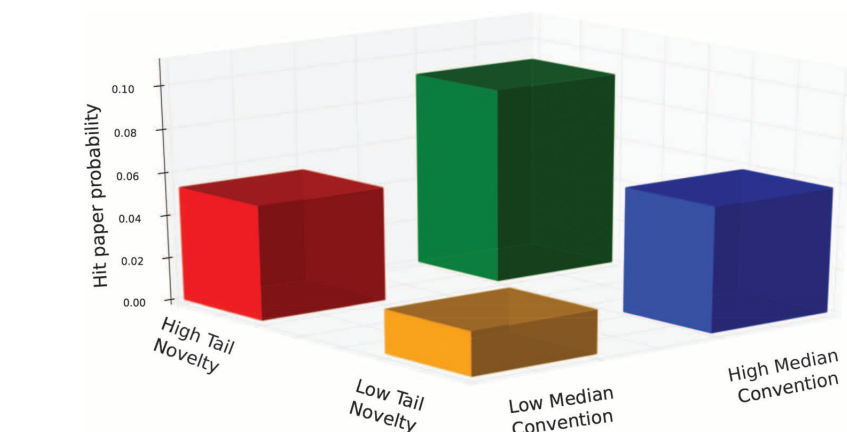

Fig. 2. The probability of a "hit" paper, conditional on novelty and conventionality. This figure presents the probability of a paper being in the top 5% of the citation distribution conditional on two dimensions: whether a paper exhibits (i) high or low median conventionality and (ii) high or low tail novelty, as defined in the text. Papers that combine high median conventionality and high tail novelty are hits in 9.11 out of 100 papers, a rate nearly double the background rate of 5%. Papers that are high on one dimension only (high median conventionality or high tail novelty but not both) have hit rates about half as large. Papers with low median conventionality and low tail novelty have hit rates of only 2.05 out of 100 papers. The sample includes all papers published in the WOS from 1990 to 2000. The supplementary materials show similar findings when considering (i) all other decades from 1950 to 2000; (ii) hit papers defined as the top 1 or 10% by citations; and (iii) analyses controlling for field and other observable differences across papers, hinting at a universality of these relationships for scientific work. The difference in the hit probabilities for each category is statistically significant ($P < 0.00001$). The percentage of WOS papers in each category are: 6.7% (green bar), 23% (gold bar), 26% (red bar), and 44% (blue bar).
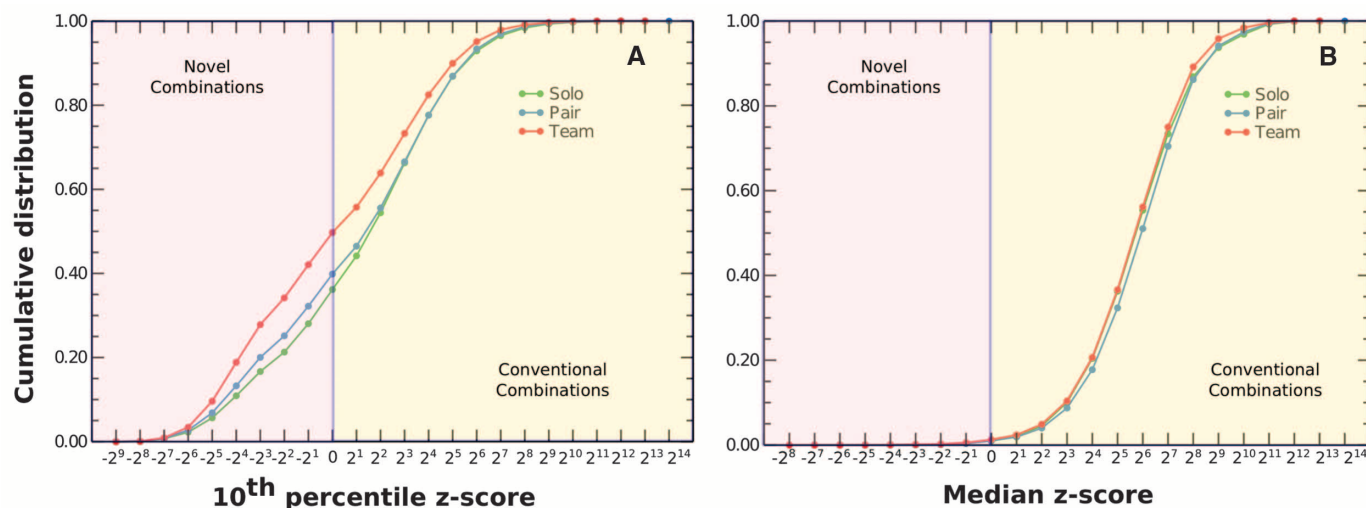

Fig. 3. Authorship structure, novelty, and conventionality. Team-authored papers are more likely to incorporate tail novelty but without sacrificing a central tendency for high conventionality. Papers introduce tail novelty (a 10th-percentile z score less than 0) in 36.2, 39.9, and 49.7% of cases for solo authors, dual authors, and three or more authors, respectively (**A**). K-S tests confirm that the distributions of tail novelty are distinct (solo versus pair $P = 0.016$, pair versus team $P = 0.001$, team versus solo $P < 0.001$). In contrast, each team size shows similar distributions for median conventionality [(**B**), K-S tests indicate no statistically significant differences]. These findings suggest that a distinguishing feature of teamwork, and teams' exceptional impact, reflects a tendency to incorporate novelty.
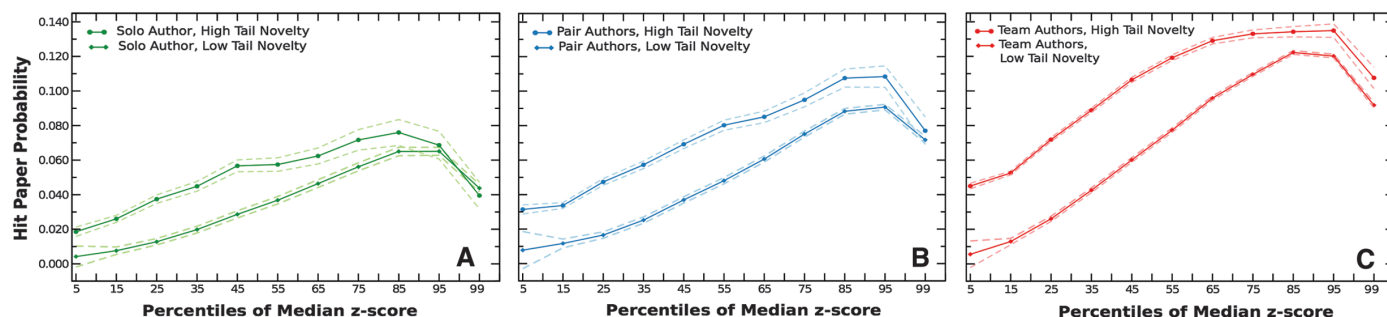
Fig. 4. Novel and conventional combinations in the production of science. (A to C) The interplay between tail novelty, median conventionality, and hit paper probabilities shows remarkable empirical regularities. First, high tail novelty papers have higher impact than low tail novelty papers at (i) any level of conventionality and (ii) regardless of authorship structure. Second, increasing median conventionality is associated with higher impact up to the 85th to 95th percentile of median conventionality, after which the relationship reverses. Third, larger teams obtain higher impact given the right mix of tail novelty and median conventionality. Nonetheless, at low levels of median convention and tail novelty, even teams have low impact, further emphasizing the fundamental relationship between novelty, conventionality, and impact in science.

Teams were neither more nor less likely than single authors or pairs of authors to display high median conventionality. Figure 3B indicates no significant statistical difference in the median z-score distributions (K-S tests indicate solo versus pair $P = 0.768$, pair versus team $P = 0.417$, team versus solo $P = 0.164$). Teams thus achieve high tail novelty more often than solo authors. Yet, teams were not simply more novel but rather displayed a propensity to incorporate high tail novelty without giving up a central tendency for high conventionality.

In our final analysis, we examined the interplay between citation, combination, and collaboration using regression methods (Fig. 4). Papers were binned into 11 equally sized categories of median conventionality. A separate regression was run for each category of median conventionality and each team size, with field fixed effects. The supplementary materials detail the regression methodology and present additional confirmatory tests (figs. S7 to S10).

There were three primary findings. First, high tail novelty papers had higher impact than low tail novelty papers, an impact advantage that occurred at any level of conventionality and regardless of authorship structure. Second, peak impact occurred in the 85th to 95th percentile of median conventionality, an exceptionally high level. This peak and its position appeared irrespective of tail novelty/no tail novelty or authorship structure. These generic features suggest fundamental underlying rules relating combinations of prior work to the highest-impact science.

Finally, Fig. 4 indicates that at virtually all mixes of tail novelty and median conventionality, larger teams were associated with higher impact. Thus, whereas teams incorporated the highest impact mixes more frequently (Fig. 3), teams also tended to obtain higher impact for any particular mix (Fig. 4). Nonetheless, despite teams' advantage in citations across virtually all fields of science (*12*), even teams had low impact at low levels of median conventionality and tail novelty.

Our analysis of 17.9 million papers across all scientific fields suggests that the highest-impact science draws on primarily highly conventional combinations of prior work, with an intrusion of combinations unlikely to have been joined together before. These patterns suggest that novelty and conventionality are not factors in opposition; rather, papers that mix high tail novelty with high median conventionality have nearly twice the propensity to be unusually highly cited.

These findings have implications for theories about creativity and scientific progress. Combinations of existing material are centerpieces in theories of creativity, whether in the arts, the sciences, or commercial innovation (*2–4, 6–10, 16*). Across the sciences, the propensity for high-impact work is sharply elevated when combinations of prior work are anchored in substantial conventionality, not novelty, while mixing in a left tail of combinations that are rarely seen together. In part, this pattern may reflect advantages to being within the mainstream of a research trajectory, where scientists are currently focused, while being distinctive in one's creativity. Combinations of prior work also relate to "burden of knowledge" theory, which emphasizes the growing knowledge demands on scientists (*4, 17, 21*). New articles indexed by the WOS now exceed 1.4 million per year across 251 fields, encouraging specialization and challenging scientists' capacity to comprehend new thinking across domains. The finding that teams preserve high conventionality yet introduce tail novelty suggests that teams help meet the challenge of the burden of knowledge by balancing domain-level depth with a capacity for atypical combinations.

This methodology considered paper and journal pairings but can be applied at the level of disciplines, papers, or topics within papers, allowing the examination of combinations of prior work at different resolutions in future studies of creativity and scientific impact. Beyond science, links between novelty and conventionality in successful innovation also appear. E-books retain page-flipping graphics to remind the reader of physical books, and blue jeans were designed with a familiar watch pocket to look like conventional trousers. From this viewpoint, the balance between extending technology with atypical combinations of prior ideas while embedding them in conventional knowledge frames may be critical to human progress in many domains. Future research questions also arise from our findings. Science is dynamic, with research areas shifting and new fields arising. Although we find that the regularities relating novelty, conventionality, and impact persist across time and fields, understanding how research trajectories shift and how new fields are born are questions that measures of novelty and convention may valuably inform. At root, our work suggests that creativity in science appears to be a nearly universal phenomenon of two extremes. At one extreme is conventionality and at the other is novelty. Curiously, notable advances in science appear most closely linked not with efforts along one boundary or the other but with efforts that reach toward both frontiers.

References and Notes

1. Committee on Facilitating Interdisciplinary Research, *Facilitating Interdisciplinary Research* (National Academies Press, Washington, DC, 2004).
2. H. S. Becker, Art Worlds (Univ. of California Press, Berkeley, CA, 1982)
3. R. Guimerà, B. Uzzi, J. Spiro, L. A. Amaral, *Science* **308**, 697–702 (2005).
4. B. Jones, *Rev. Econ. Stud.* **76**, 283–317 (2009).
5. B. F. Jones, S. Wuchty, B. Uzzi, *Science* **322**, 1259–1262 (2008).
6. J. Schumpeter, *Business Cycles* (McGraw-Hill, New York, 1939).
7. A. P. Usher, *A History of Mechanical Invention* (Harvard Univ. Press, Cambridge, MA, 1954).
8. M. L. Weitzman, *Q. J. Econ.* **113**, 331–360 (1998).
9. M. Schilling, *Creat. Res. J.* **17**, 131–154 (2005).
10. B. Uzzi, J. Spiro, *Am. J. Sociol.* **111**, 447–504 (2005).
11. H. J. Falk-Krzesinski et al., *Res. Eval.* **20**, 145–158 (2011).
12. S. Wuchty, B. F. Jones, B. Uzzi, *Science* **316**, 1036–1039 (2007).
13. D. Stokols, K. L. Hall, B. K. Taylor, R. P. Moser, *Am. J. Prev. Med.* **35** (suppl.), S77–S89 (2008).
14. S. M. Fiore, *Small Group Res.* **39**, 251–277 (2008).
15. J. A. Evans, J. G. Foster, *Science* **331**, 721–725 (2011).
16. R. Collins, *The Sociology of Philosophies: A Global Theory of Intellectual Change* (Harvard Univ. Press, Cambridge MA, 1998).
17. L. Fleming, *Manage. Sci.* **47**, 117–132 (2001).
18. M. Schilling, E. Green, *Res. Policy* **40**, 1321–1331 (2011).
19. R. M. Henderson, K. B. Clark, *Admin. Sci. Q.* **35**, 9–30 (1990).

20. P. Azoulay, J. G. Zivin, G. Manso, *Rand J. Econ.* **42**, 527–554 (2011).
21. A. Einstein, *The World as I See It* (Citadel Press, Secaucus NJ, 1949).
22. D. T. Whiteside, *J. Hist. Astron.* **1**, 116–138 (1970).
23. H. Small, *J. Am. Soc. Inf. Sci.* **24**, 265–269 (1973).
24. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, *J. Am. Soc. Inf. Sci. Technol.* **61**, 1377–1385 (2010).
25. M. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, *PLOS One* **3**, e1683 (2008).
26. S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, U. Alon, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**, 026127 (2003).

A Radical Intermediate in Tyrosine Scission to the CO and CN⁻ Ligands of FeFe Hydrogenase

Jon M. Kuchenreuther,[1]* William K. Myers,[1]* Troy A. Stich,[1] Simon J. George,[1] Yaser NejatyJahromy,[1] James R. Swartz,[2] R. David Britt[1]†

The radical *S*-adenosylmethionine (SAM) enzyme HydG lyses free L-tyrosine to produce CO and CN⁻ for the assembly of the catalytic H cluster of FeFe hydrogenase. We used electron paramagnetic resonance spectroscopy to detect and characterize HydG reaction intermediates generated with a set of ^{2}H, ^{13}C, and ^{15}N nuclear spin-labeled tyrosine substrates. We propose a detailed reaction mechanism in which the radical SAM reaction, initiated at an N-terminal 4Fe-4S cluster, generates a tyrosine radical bound to a C-terminal 4Fe-4S cluster. Heterolytic cleavage of this tyrosine radical at the C_α-C_β bond forms a transient 4-oxidobenzyl (4OB•) radical and a dehydroglycine bound to the C-terminal 4Fe-4S cluster. Electron and proton transfer to this 4OB• radical forms *p*-cresol, with the conversion of this dehydroglycine ligand to Fe-bound CO and CN⁻, a key intermediate in the assembly of the 2Fe subunit of the H cluster.

Microbial hydrogenase enzymes catalyze the redox interconversion of protons and H_2 by using earth-abundant metals in their catalytic centers, with NiFe, Fe, and FeFe classes known (*1*). The FeFe hydrogenases are adept H_2 producers, with turnover frequencies up to 10,000 s^{-1} (*2*). Their catalytic H cluster (Fig. 1) consists of a conventional 4Fe-4S cluster linked to a unique 2Fe cluster that has two CN⁻ ligands, three CO ligands, and a dithiolate bridge with a central atom X, the chemical identity of which is ambiguous in current x-ray structures (*3, 4*). However, a nitrogen assignment to atom X is supported by recent reports of the assembly of active FeFe hydrogenase by incorporation of a synthetic 2Fe subcluster with an azadithiolate bridge into apoenzyme (*5, 6*).

The HydE, HydF, and HydG maturases are Fe-S cluster–containing accessory proteins involved in the biological synthesis of the 2Fe component of the H cluster (*7*). HydE and HydG are members of the ever-growing family of radical *S*-adenosylmethionine (SAM) enzymes, characterized by a CysX$_3$CysX$_2$Cys motif, where the three

cysteine residues coordinate a 4Fe-4S cluster through ligation of three Fe ions. The fourth Fe of the cluster binds SAM as a N/O chelate, and SAM is reductively cleaved to produce methionine plus a strongly oxidizing 5-deoxyadenosyl (5′-dA•) radical (*8, 9*). A diverse array of reactions is powered by the H-atom abstraction capability of the resultant 5′-dA•, and current bioinformatics surveys reveal almost 50,000 members of the radical SAM enzyme class (*10*).

The focus of this report is the HydG maturase. This radical SAM enzyme generates CO, CN⁻, and *p*-cresol by using free tyrosine as its substrate (*11–14*). Although HydG has yet to be crystallographically characterized, biophysical data indicate that it has two 4Fe-4S clusters (*15, 16*). The SAM–4Fe-4S cluster is bound near the N terminus, whereas the second cluster is modeled as coordinated to three cysteine residues of a CysX$_2$CysX$_{22}$Cys sequence near the C terminus. Sequence homology with other radical SAM enzymes, such as biotin synthase and the tyrosine lyase ThiH, point to HydG having a triose phosphate isomerase barrel structure, in which eight α helices surround a ring composed of eight parallel β strands that contain a buried active site suitable for small-molecule substrates (*11, 17*).

Like HydG, ThiH also lyses tyrosine, in this case to generate dehydroglycine (DHG) as an intermediate in anaerobic thiamine biosynthesis

(*18, 19*). ThiH lacks the second 4Fe-4S binding domain of HydG. The electron paramagnetic resonance (EPR) spectrum of the ThiH SAM-[4Fe-4S]⁺ cluster is altered by SAM binding but unaffected by tyrosine incubation. In a proposed mechanism (*18, 19*), the initial 5′-dA• abstracts the phenolic H of the tyrosine substrate, forming a neutral tyrosine radical, which is then cleaved at the C_α-C_β bond. Quantum chemistry calculations favored homolytic cleavage of this bond to form a transient glycyl radical, given its lower energy pathway compared with heterolytic cleavage (*19*). Modeling the differences in reactivity between wild-type HydG and a HydG mutant missing the C-terminal Fe-S cluster (HydG^{SxxS}), which produces some CN⁻ but no CO, led Nicolet and co-workers (*20*) to propose the same glycyl radical intermediate in the HydG mechanism, building on the thermodynamic argument for the mechanism in ThiH. However, no radical intermediates have been experimentally characterized for either enzyme.

Here, we report on the EPR spectroscopy of wild-type *Shewanella oneidensis* HydG (HydG^{WT}) expressed in *Escherichia coli*. Such expressed HydG, combined with HydE and HydF, can be used for in vitro synthesis of the 2Fe component of the H cluster and concurrent activation of FeFe hydrogenase apoprotein (*14, 21*). The use

[1]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA. [2]Department of Chemical Engineering and Department of Bioengineering, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
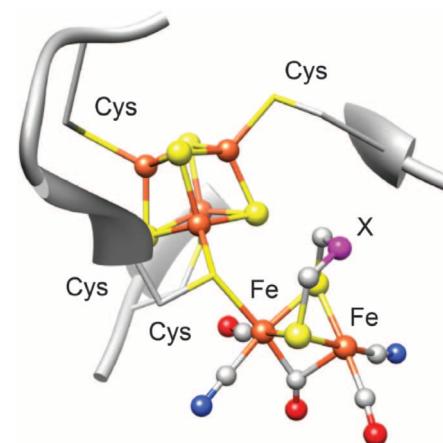†Corresponding author. E-mail: rdbritt@ucdavis.edu

Fig. 1. The catalytic H cluster of FeFe hydrogenases. Ball-and-stick representation (from Protein Data Bank entry 3CY8) was generated by using UCSF Chimera (www.cgl.ucsf.edu/chimera/): Fe (brown), S (yellow), C (gray), O (red), N (blue), and atom X (magenta). H is not shown for simplicity.