Volume Title: The Changing Frontier: Rethinking Science and Innovation Policy

Volume Author/Editor: Adam B. Jaffe and Benjamin F. Jones, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 0-226-28672-X, 978-0-226-28672-3

Volume URL: http://www.nber.org/books/jaff13-1

Conference Date: August 2-3, 2013

Publication Date: July 2015

Chapter Title: Comment on "Collaboration, Stars, and the Changing Organization of Science: Evidence from Evolutionary Biology"

Chapter Author(s): Julia Lane

Chapter URL: http://www.nber.org/chapters/c13039

Chapter pages in book: (p. 102 – 106)

Romer, P. M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–102.
Rosenbloom, R. S., and W. J. Spencer. 1996. *Engines of Innovation: US Industrial Research at the End of an Era*. Boston: Harvard Business Press.
Tang, L., and J. P. Walsh. 2010. "Bibliometric Fingerprints: Name Disambiguation Based on Approximate Structure Equivalence of Cognitive Maps." *Scientometrics* 84 (3): 763–84.
Waldinger, F. 2012. "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany." *Review of Economic Studies* 79 (2): 838–61.
———. 2013. "Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge." Working Paper, University of Warwick.
Weitzman, M. L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113 (2): 331–60.
Wuchty, S., B. F. Jones, and B. Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–39.

Comment Julia Lane

The authors address an interesting and important question about the way in which scientific collaboration has changed over time. They use a creatively constructed data set on evolutionary biology to show how scientific collaboration for the subset of authors they identify has changed. The results reported are consistent with other work in the book. Most interestingly, the geographic distance between coauthors has increased substantially, notably that the concentration of publications within an institution has decreased and that the institutional rank distance between coauthors has increased. They find that the concentration of publications at the individual level has increased. They also note the pool of potential coauthors has increased. The authors posit that these trends are the result of two factors: the burden of knowledge and collaboration-supporting technologies.

Their work thus provides potential new areas that could be examined in future research. One is whether evolutionary biology is unique among scientific disciplines: it would be extremely useful to know whether similar changes in collaboration are found for such subsets of authors in "big science," like astrophysics, and smaller scale sciences, like chemistry. It would also be useful to examine across different disciplines whether observed changes in collaboration are due to specialization of innovative labor. It would also be very useful to understand the role of technology in driving geographically dispersed collaboration. It is possible, and anecdotal evidence suggests, that increasingly technology-intensive science that requires large-scale complex equipment is a driving force behind changes in collaboration.

Julia Lane is an Institute Fellow at the American Institutes for Research.

For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see http://www.nber.org/chapters/c13039.ack.

The research also raises interesting questions about whether a field, even though it might be named the same over time, is indeed still the same. Does evolutionary biology still mean the same thing now as it did twenty-nine years ago? Or has it now become more interdisciplinary and been influenced by the convergence of physics, chemistry, and biology? In other words, how much are changes in collaboration due to changes in the very structure of science itself? There is a great deal of interest in using natural language processing techniques to study the evolution of scientific disciplines; future research could examine the role of changes in the topics covered under different disciplinary nomenclatures on collaboration (Herr 2009; Talley et al. 2011).

The findings also suggest interesting potential research into other areas—particularly the role of monetary incentives in changing collaboration. Over the period that is studied, US science agencies changed the way in which science was funded. Although the big interdisciplinary funding initiatives—such as funding for nanotechnology and the human genome sequence—are well known and fundamentally changed both the nature and scale of scientific endeavor in a number of fields, there have been many such smaller scale initiatives. In addition, during the period of study, funding abruptly doubled for the NIH and then leveled off—and so the sharp change in funding could be used to examine the role of monetary incentives.

The authors also posit some interesting hypotheses about the role of graduate students and postdoctoral fellows in the changes in collaborations they observe among their subset of authors. We know that the composition and size of scientific teams is not only dramatically different across research fields, but also changes substantially in response to monetary incentives (Stephan 2007, 2012). Subsequent research on the role of team structure in collaboration could examine both how the structure of teams in evolutionary biology changed over the period—and whether and how subsequent placements of team members in industry or in academia evolved over the time period in question.

Much can also be learned from the extensive hard work that was done to structure the data set. The chapter is partly about authors and partly about institutions, which makes it very difficult to develop a representative frame. The authors have done an enormous amount of work to develop a frame that is based on a very specific selection of publications; this is the link asset that includes information about both.

Future research could be very useful to inform us about the generalizability of the authors' results. In particular, it would be extremely useful to learn whether the known skewness of productivity of scientists, of institutions, and of the salience of publications yields an analysis that is representative of all three dimensions. There has been a rich vein of research in other fields, like labor economics, to understand how the link between workers and firms (jobs) can become a frame of study in its own right (Abowd and Vilhuber 2011; Abowd, Haltiwanger, and Lane 2004; Burgess, Lane, and Stevens 2000);

similar research could usefully be undertaken using data sets constructed in the way described in this chapter. I outline a possible research agenda below.

The Link Asset: Publications. The authors use an imaginative approach to construct a frame from which to draw their sample. In particular, the frame described in this chapter is derived from publications in four journals over twenty-nine years whose focus is evolutionary biology. It would be extremely useful for other researchers to determine whether frames constructed like this are representative of research fields. Are such frames constant over time, both with respect to coverage and content? In particular, how does the emergence of new journals that might siphon off contributors affect the representativeness of the frame? For example, in economics, would a frame based on the publications *Journal of Political Economy*, *Quarterly Journal of Economics*, *American Economic Review*, and *Econometrics*, for example, be equally representative over time, both in levels, and as a proportion of all journals in which economists publish? New research in computer science is beginning to enable us to determine what is missing—by using the same capture/recapture techniques that are used to generate censuses of wildlife, we can determine the coverage of journals over time (Khabsa, Treeratpitu, and Giles 2012; Giles and Khabsa 2014).

The authors use an equally imaginative approach to weighting their sample, and their approach should stimulate an interesting line of research. In particular, they use citations to weight publications, rather than the absolute measure of publications themselves. Returning to my earlier analogy about what we have learned from researching firms, if the focus of the research is firms, then firms can be used as the unit of analysis. If the focus of research is employment, then it is more appropriate to use employment-weighted firms. In the former case, the small start-up has the same weighting as does General Motors. In the latter case, General Motors' behavior will dominate the analysis. Similarly, in this chapter, the results will be dominated by the activities of a few highly productive (and cited) researchers; an important line of analysis for future research would be to examine how the results change when all individual researchers are weighted equally.

The Article Authors. The sample design used in the chapter suggests even more useful research that could be undertaken to build a richer understanding of how the empirical approach can be expanded to understand the activities of active researchers. For example, the original draw from the four journals identifies 171,428 authors; 140,240 of these are dropped because there are no more than two publications linked to their names. Additional research could examine the sensitivity of the results to, for example, distance measures (how many authors are dropped because it is not possible to identify nonunique author names, and understanding the selection bias associated with nonunique Asian author names); time variability (how does the sample selection change over time); and institutional variation (how many institutions are dropped). There are equally interesting questions about the impor-

tance of using fractionally weighted output measures rather than full output measures. In the case of this chapter, the authors decide to weight author output measures by the full publications; since coauthorship has increased overtime, this decision weights output in later time periods more heavily than previous time periods. Much useful research could examine how different weighting choices affect the econometric results. At the least, both weighted and unweighted results should be presented and discussed throughout.

The chapter also identifies an important possible set of sources for collaboration—namely the possibility of faculty coauthoring papers with their postdoctoral fellow and graduate students. This builds on a rich literature that has suggested that within-group specialization is an important feature of modern science (Black and Stephan 2010; Conti, Denas, and Visentin, forthcoming). It would be extraordinarily valuable to build a database that included information about the links between the authors (including the postdoc/graduate student/principal investigator relationship) to establish this hypothesis more generally.

The Institutions. Another important potential line of research that is identified in this chapter is the role of institutional locations, about which so little is known. The structure of the sample identifies some interesting features of collaboration. Just over half (57 percent) of the papers have a single institution listed, and so all authors are located with this institution, and 79 percent of authors are attributed to an institution. Future research could examine how a sample constructed like this might influence our understanding of the role of institutions—particularly focusing on how the structure of institutional affiliations changes over time, both domestically and internationally—and how a selection decision based on institutional affiliations affects the number of coauthors included in the analytical data set.

Finally, the authors put together a very interesting model about the role of graduate and postdoctoral students in generating increasing collaborations. This framework can be used to add to the three alternative hypotheses that are also bruited in the literature. In principle, an empirical exercise could test the model by including data on the pool of available graduate students, data on the role of funding agencies in incentivizing collaboration, the role of technological complementarity, as well as the role of previous collocation.

In sum, this chapter presents some very provocative results, which are very congruent with companion chapters in the book. The challenge to the research community is to extend their interesting results generated from a painstakingly assembled and idiosyncratic sample and determine the generalizability of the approach to both evolutionary biology and other areas of science.

References

Abowd, J., J. Haltiwanger, and J. Lane. 2004. "Integrated Longitudinal Employee -Employer Data for the United States." *American Economic Review* 94:224–29.

Abowd, J. M., and L. Vilhuber. 2011. "National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail." *Journal of Econometrics* 161:82–99.

Black, G., and P. E. Stephan. 2010. "The Economics of University Science and the Role of Foreign Graduate Students and Postdoctoral Scholars." In *American Universities in a Global Market*, edited by C. Clotfelter, 129–61. Chicago: University of Chicago Press.

Burgess, S., J. Lane, and D. Stevens. 2000. "Job Flows, Worker Flows, and Churning." *Journal of Labor Economics* 18 (3): 473–502.

Conti, A., O. Denas, and F. Visentin. Forthcoming. "Knowledge Specialization in PhD Student Groups." *IEEE Transactions on Engineering Management.* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2210462.

Giles, L., and M. Khabsa. 2014. "The Number of Scholarly Documents on the Web." *PLoS One.* doi: 10.1371/journal.pone.0093949.

Herr, B. W., E. M. Talley, G. A. P. C. Burns, D. Newman, and G. LaRowe. 2009. "The NIH Visual Browser: An Interactive Visualization of Biomedical Research." Information Visualization, 13th Annual Conference, Barcelona, July. Institute of Electronics and Electronic Engineers. http://ieeexplore.ieee.org/xpl/abstractAuthors.jsp?tp=&arnumber=5190811&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5190811.

Khabsa, M., P. Treeratpitu, and L. Giles. 2012. "AckSeer: A Repository and Search Engine for Automatically Extracted Acknowledgments from Digital Libraries." Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. http://clgiles.ist.psu.edu/pubs/JCDL2012-AckSeer.pdf.

Stephan, P. 2007. "Wrapping It Up in a Person: The Mobility Patterns of New PhDs." In *Innovation Policy and the Economy*, vol. 7, edited by A. Jaffe, J. Lerner, and S. Stern. Cambridge, MA: MIT Press.

———. 2012. *How Economics Shapes Science.* Cambridge, MA: Harvard University Press.

Talley, Edmund M., David Newman, David Mimno, Bruce W. Herr II, Hanna M. Wallach, Gully Burns, Miriam Leenders, and Andrew McCallum. 2011. "Database of NIH Grants Using Machine-Learned Categories and Graphical Clustering." *Nature Methods* 8:443–44. doi:10.1038/nmeth.1619.