Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 52, No. 3 (2001) 365–377

Feature report

Reflections on scientific collaboration (and its study): past, present, and future*

DONALD DEB. BEAVER

Bronfman Science Center, Williamstown, MA (USA)

Personal observations and reflections on scientific collaboration and its study, past, present, and future, containing new material on motives for collaboration, and on some of its salient features. Continuing methodological problems are singled out, together with suggestions for future research.

Introduction

Derek J. deSolla Price, Eugene Garfield, Henry Small, and Belver Griffith, among others, the real pioneers of the systematic study of collaboration in scientific research, as well as early and fundamental contributors to the creation of scientometrics, have left a lasting legacy. Forty years after their groundbreaking work, a large and growing number of scholars spanning the globe and four continents follow in their footsteps, extending and expanding what we know about the structure and dynamics of collaboration.

In particular, it is significant to have so many researchers at work in China and India, representing a third of humanity, and, presumably eventually a third of all scientific and technological research. It is a truism in the history of science and technology that no one region, nation, or civilization remains the center of creativity and activity for long. One need only think of the historical path of science through Mesopotamia, Greece, Islam, the Medieval Latin West, Northern Europe, the United States and Soviet Union, to grasp the point.

* Keynote speech presented at the Second Berlin Workshop on Scientometrics and Informetrics / Collaboration in Science and Technology and First COLLNET Meeting, Sept. 1, 2000, Neuen Hohendorf, Germany

0138–9130/2001/US $ 15.00

In that regard, we stand at the beginning of what appears to be another important transition in the leadership of science and technology, in the history of civilization. An international view is even more important than before, because the world as a whole, and the research world of science and technology with it, is undergoing a major transformation, the exact dimensions of whose nature and future are not yet clear, and may not be for decades. As globalization and internationalization continue, on the way to the formation of a global community, emphasis on cooperation and group life become an increasingly common counterpoint to an existing emphasis on competition and individuality. What the eventual balance will be, or should be, is not ours to tell, even though the change involves the familiar age-old problem of finding a balance between the individual and society.

Situated as we are then, in the midst of an important transitional period, it is appropriate to take stock of the past and the changing present, to reflect upon the study of scientific collaboration.

Structure

The following remarks offer a series of personal observations and reflections on scientific collaboration and its study, past and present, and make a few tentative observations about the future (not many, because the future is so uncertain). Occasionally, I hope to single out areas where there are continuing methodological problems, as well as even suggest future areas for research. What follows falls into three parts:

1. The Past
   a. A review of Beaver and Rosen, 1978-79;
   b. Teamwork (Big Science) differs from collaboration (little science);
   c. Changes in collaboration resulting from changes in research organization.
2. The Present
   d. Collaboration from the researchers' viewpoint(s).
3. The Future
   e. Remarks on email, and the world wide web.

Studies in Scientific Collaboration, [1978]

Using bullet points, let me briefly summarize the chief unusual or novel findings of the 1978 papers, published by Richard Rosen* and me (*Beaver* and *Rosen*, 1978; 1979)

- Collaboration was almost exclusively by French chemists in the period 1800–1830.
- Collaboration grew slowly until World War I, after which it grew at a much more rapid rate.
- The statistics of collaborative authorships follow a Poisson distribution, signifying a relatively rare event, gradually tending to a negative binomial distribution as collaboration became more frequent.
- The MODE of coauthorship was 2. (It still is today, especially if one counts laboratories instead of individual coauthors.)
- A collaborative first paper meant later above average productivity.
- Core journals have higher frequencies of collaborative papers than the average journal.

This last point is the basis for an important caution about research methodology in studies of scientific collaboration. Although the simplest procedure for obtaining a data sample is to use the ISI database, or to do a select sample of a few core journals, such journals are unrepresentative of the whole. Scientists themselves are generally unaware of the differences among journals, taking as their models the key journals in their fields. Core journals form a small yet visible elite, and, as such, display characteristics of the scientific elite, which may be several generations in advance of the whole of science, speaking socioculturally about research practice. Looking primarily at core and prestigious data sources will bias one's picture; studies concentrating on such data need to qualify their results accordingly.

From collaboration to TEAMWORK [1984] (Beaver, 1984)

- Discontinuity in the organization of scientific research: from little science to Big Science, ca. WWII.
- Teamwork, or giant collaborations multiply after WWII: high energy physics (HEP) is the exemplar.

---

* Richard Rosen was a student of mine in the late 60s who went on to study with Robert K. Merton at Columbia University, receiving a master's degree in sociology. Today he lives in New York City with his family, and is self-employed, no longer in academia.

- Coauthorships in giant collaborations (teamwork) follow a power law distribution, different from the Poisson characteristic of "traditional" small collaborations.
- There is no simple distribution making the coauthorship distributions of *teamwork* continuous with those of small ($N \leq 5$) *collaborations*. Whether a general distribution exists remains a puzzle.

Speaking of statistical puzzles, one of the puzzling statistical features of communication in the sciences is one noted in the 1960s that to a first approximation, as measured scientometrically, formal communications amongst scientists are random. That research indicated that the Signal to Noise (S/N) ratio in scientific communication was very small (the almost universal complaint of scientists that most of the literature is garbage may seem to confirm that finding). But we might extend that research to collaboration insofar as it reflects communication in science. Then, presumably there, too, the majority of collaborative relationships are also random. Yet it is clear that at the individual level, collaborations and communications are made with purpose and intention. How is it possible to produce such randomness out of so many purposeful, (one might even say causally related) decisions to communicate or collaborate? A satisfactory resolution of the puzzle might have important implications for the analysis and description of science, and of other social structures in which an apparently high degree of stability and order is maintained by a relatively small set of practices.

- Teamwork, or giant collaborations, represents a *new paradigm* for the organizational structure of research.
- Teamwork has spread from HEP, most notably to molecular biology and biomedical research. See, for example, the human genome project (HUGO).

The changing organizational structure of research

Over the past few years, Henry Etzkowitz, among others, has been gradually constructing a new view of the organization of scientific research more consistent with Big Science, in which the research scientist plays the role of entrepreneur. Because the research carried out in such a style of scientific organization is almost wholly collaborative, the implications of how that organization is implemented in the laboratory are directly relevant to undestanding collaboration in research. What follows briefly outlines the advantages and disadvantages of that organization, both as reflected in Etkowitz' work, and as supplemented through interviews with some of my scientist colleagues.

The typical group *structure* at a major research university consists of: A Principal Investigator (PI), together with postdocs, graduate students, (and perhaps undergraduates) -or- A senior professor, perhaps an assistant or junior professor, postdocs, graduate students, (and perhaps undergraduates).

Salient Advantages:

Efficiency, Power "Many hands make light work."

- Multiplicity of projects optimizes chances for funding, for obtaining support for the lab and continuing research.
- "A stable of graduate students is a power booster."*

Speed

- Like the advantages, in some cases, of parallel processing. Can parcel out parts of a problem, and finish more rapidly than one's competition.
- Students are already trained, OR, the seniors train the juniors. Lab leader freed from the time it takes to train new researchers.

Breadth:

- Can tackle broader, more encompassing problems, "more exciting things." Consequently enhances visibility and feedback at meetings.

For example, paraphrasing a geologist at Williams College, "I can put one student into the field for the summer, 3 months. After 5 years, I'll have enough data to produce a research publication. A large research group can put 5 students in the field for the summer, 3 months. But in 3 months, the research group already has the data for a publication."**

Synergy

- Multiplicity of viewpoints energizes and excites participants. Makes actual work more intense.

Reduced Risk ["Don't place all your eggs in one basket."]

- Can have several projects going simultaneously; increases chances of success, and successful re-funding.

---

* Science Professor, Williams College, private communication, August, 2000.

** Science Professor, Williams College, private communication, August, 2000.

Flexibility maintained

- Can have one project of a “far-out,” speculative, and prospective nature.
- Failure does not destroy the laboratory. Success may open up new directions, funding sources that accrue to pioneer leaders of new successful research program.

Accuracy

- Errors are more readily detected when several different individuals with different perspectives discuss or argue about data and/or theory. Another way to view this is that in collaboration, the “context of justification” becomes to some extent part of the “context of discovery”, or that a large collaborative group partly embodies the valuable and ongoing process of intersubjective verifiability.

Feedback, Dissemination, Recognition and Visibility

- Participants can present preliminary findings at many different colloquia or conferences and get response from their colleagues. They can more widely disseminate their findings, and lay claim to their piece of the research turf.

Disadvantages

Individuals’ invisibility

- Most participants are invisible, in a formal sense, to the larger research community. They are just “names” on a paper, “fractional” scientists, essentially anonymous.

PI loses touch with direct research

- Reduces creativity inspired by directly acquired tacit knowledge of how things work in practice.
- Loses ability to be a bench scientist.
- Diverts creative talents to administration, competition for limited resources, rather than actual research.

Privatization of Research harmful to research ethos

- Creation of entrepreneurial fiefdoms may promote tempting negative strategies, especially secrecy or additional limits on the free sharing of ideas and materials in research.
- Cooperation with other laboratories (competitors) may be for purposes of cooptation or espionage, practices potentially harmful to science. Even if for the more positive purpose of alliance, competitive advantage may deter “smaller” laboratories or individuals.

It is an open question whether and how such an organizational style can long continue, given individual's self-interest in obtaining recognition of their own creativity.

Note that viewing collaboration primarily from a laboratory perspective creates another interpretive possibility for understanding collaborative work: Collaborations of 10–12 people could be viewed as another level of the original historical Poisson-type collaborations: Two different research groups, each of size 5 to 6, led by a PI, collaborate. Each research group could be seen as a kind of "person", or "individual", just as in (American) law, a corporation as a legal entity is a person or being. Then, such collaborations are really 2-"author" collaborations, in which the individual human researchers are but component parts of larger wholes. Being a "component" may be satisfactory through the postdoctoral years, for security and acquisition of new skills, but thereafter, the ambitious individual will want to become a PI.

By this interpretation we have a kind of hybrid collaboration lying between "collaboration" and "teamwork." Having 10–12 individuals working on the same project should qualify their product as teamwork, but if they are viewed as 2 collective individuals (laboratory collectives), their product is like old style collaboration. The fact that the modal number of collaborating laboratories is 2 additionally supports this idea of laboratories/working groups as "individuals." Furthermore, this relatively new way of organizing research fits and extends nicely Derek Price's suggestion that collaboration is in part a response to a shortage of scientists, allowing there to be "fractional" scientists. (*Price* and *Beaver*, 1966)

Research scientists' views on collaboration today. Background

The following comments reflect the views of currently active researchers about what collaboration means to them, based on a series of one-hour long interviews.* Their perspectives on collaboration derive from the standpoint of an elite United States' liberal arts college, located in Williamstown, in Northwestern Massachusetts.

Williams College is a coeducational undergraduate college of about 2,000 eighteen to twenty-one year olds, about evenly split between male and female students. It is very highly rated academically and it students are on a par with those of major research universities like Harvard, Yale, Princeton, Berkeley, and Stanford for admission. About 40% of the students major in the natural sciences, mathematics, computer science, and psychology. Williams leads small college in terms of National Science Foundation Grants per science faculty member.

* In all there were 7 scientists: 2 computer scientists, 2 physicists, 1 geologist, 1 biologist, and 1 chemist.

For shedding light on collaboration, Williams has the following 3 advantages:

[1] (*reproduces researchers*) Small liberal arts colleges are "feeders," to science: per capita undergraduate student, they lead to more Ph. Ds in science than major research universities, and that has been true for most of the 20th century. (See *Knapp* and *Goodrich*, 1952.)

[2] (*"hands-on" learning by doing collaborative research*) A great educational advantage of the small liberal arts college is that undergraduates actively participate in on-going research front investigations. They do so both because pedagogically such experience affords superior education, and because their mentors reciprocally derive benefit from their activity in the laboratory. Many undergraduate students publish their first research paper with their advisers; a significant fraction of professors' publications consists of paper written with student co-authors.

[3] **(***clearer standpoint*) Over the past few decades, pressures for greater research productivity at liberal arts colleges has increased, to the point where researchers at such institutions compare with those at minor research universities. Thus being active in research, but not in a major research university, research institute, or industrial research lab, affords a unique vantage point for providing a clearer perspective on the nature and function of collaboration. It is to be hoped that such a standpoint may help correct or make more objective findings based only upon data from the most elite major research institutions.

Perspectives on collaboration

Let us proceed then, to see what my colleagues said about motives for collaboration in research – why do they do it? First let us consider the summary outline presented in Table 1.[*]

In large measure, the summary items presented in Table 1 speak for themselves, so I won't dwell on them here, except to emphasize the very welcome item number 18 – if we ever lose sight of those motives, we're in trouble. There are, however, some additional significant themes that emerged in response to five other questions about collaboration.

[*] For a related table, dealing with 10 general factors helping to increase collaboration, see *Katz* and *Martin*, 1997, Section 2.2.

Table 1
The purposes for which people collaborate

1	Access to expertise.
2	Access to equipment, resources, or “stuff” one doesn’t have.
3	Improve access to funds.
4	To obtain prestige or visibility; for professional advancement.
5	Efficiency: multiplies hands and minds; easier to learn the tacit knowledge that goes with a technique.
6	To make progress more rapidly.
7	To tackle “bigger” problems (more important, more comprehensive, more difficult, global).
8	To enhance productivity.
9	To get to know people, to create a network, like an “invisible college”.
10	To retool, learn new skills or techniques, usually to break into a new field, subfield, or problem.
11	To satisfy curiosity, intellectual interest.
12	To share the excitement of an area with other people.
13	To find flaws more efficiently, reduce errors and mistakes.
14	To keep one more focussed on research, because others are counting on one to do so.
15	To reduce isolation, and to recharge one’s energy and excitement.
16	To educate (a student, graduate student, or, oneself).
17	To advance knowledge and learning.
18	For fun, amusement, and pleasure.

How do collaborations start?

- By chance, at a colloquium or lecture, or at a conference, because of a presentation, or because of working sessions or, on leave at another institution, to learn new skills, or catch up with the field.
- By intention, by letter or phone call of solicitation.
- By recommendation or referral by trusted colleagues.
- Because it’s a part of one’s job – to mentor, to educate.

What’s the typical size of a collaboration?

- 2 or 3 persons or laboratories, OR “giant”.
- Dominant model: 2 individuals, usually “peers”.
- Unusual persistence of “Poisson” model, of *pairing* off.
- Perhaps also responds to the pressure of unwanted intermediate authorships – with 2 authors, can take turns at being first author.

- Persistence of prestige of single-author publications (perhaps also dependent on the journal where published).
- Some even “frown on” collaborations of more than 6 people.
- It shows “you still have the juice to do it on your own”.
- This suggests that in our studies of collaboration, we should also pay more attention to single authors, as counterpoint, balance, and for comparative purposes to help calibrate and place our results in context.

How is credit allocated in collaborations?

- Name Ordering: a signal to the research community, and to hiring committees’ evaluations, which at Wiliams often first look at the total publication list, then look for the percentage of first author, or, single authored papers, as a sign of creative independence and ability to do most of the work of a published piece of research – qualities needed to establish a research laboratory, get funding, and educate students in the laboratory.
- Conventions are highly variable, and dependent on field or subfield. Alphabetical or First Place-Last Place are the two most common systems. Conventions vary enormously. Intermediate authors tend to be overlooked, or, intermediate authorships tend to be less highly valued.
- A rather unique way of determining authorships is practiced by a theoretical quantum information group at IBM, where the group leader lists everyone who contributed to the research, and then invites individuals who don’t feel they did enough to deserve an authorship to cross their names off the list.

Does collaboration affect one’s research productivity?

- At worst, collaboration doesn’t influence, at best, it enhances.
- Problems: The persistence of stylistic differences complicates evaluation. For example, consider the different practices represented by the following types of research practice: field-closet; field-lab; theoretical-experimental. Furthermore, subtle but significant differences in co-authorships and also the frequency of collaboration

may be lost or simply undetectable in aggregate data. It is important to know something qualitatively about the nature of the research being studied, and who is performing it.

Has Email affected collaborations?

- Generally, research is nearly impossible without it.
- Cf. one scientist's collaboration with colleagues from China, Russia, and Mexico.
- Cf. one scientist's collaboration with colleagues at 3 different universities (e.g. California, Munster, Bremen).
- Enhances efficiency, intensiveness, if not necessarily volume – for some. But others clearly wouldn't be as productive w/o email-assisted collaboration.

The future: Internet and E-journals

There is space and time for only a few limited and necessarily speculative ideas about possible future changes that may affect the form, quality, and nature of collaborative research in the future. In particular, the expansion of the World Wide Web, and the growing number of electronic journals are likely to bring changes in research practice, which will be in turn reflected in the conventions of formal "publication", whether singly or multiply-authored.

Because science is "many-brained", as Derek Price used to like to say, the open and accessible nature of sites and links on the Web is tailor made to suit that character. But, just as important as surfing or searching for data may be, it is equally important to know when to stop doing so.

Because data is becoming so ubiquitous, and web sites proliferating, the practice of taking people's materials off the web and manipulating them for research, for lectures, or other professional purposes, is bound to increase. There will be enormous temptation to do "instant research." With increasing "borrowing" of others' materials will come problems of determining, assuring, or evaluating quality. (At least a few of our undergraduates are already adept at relying on the Web for research papers, while still neophytes at evaluating the validity or adequacy of the data they acquire.) Because the Web simultaneously becomes both investigative tool and research subject, how to deal with that novel character will require considerable care.

For collaborations, and collaborative study in particular, increasing globalisation is likely to produce increases in the geographical diversity of collaborators, be they individuals, laboratories, or institutes. Physical location is no longer a barrier to the free and easy exchange of information. Indeed, it may be the case that the advent of email had already begun to increase diversity in geographical locations. It would be interesting to see if such a phenomenon could be detectable in a retrospective study.

Collaborative research published in ejournals will, for a while, create enormous problems of comparison with that represented by print journals, and quite likely many of the problems that arise in evaluating the latter will also apply to the former. It is not yet clear what will constitute the "core" of ejournals, or along what lines they will be stratified. Perhaps the simplest strategy for evaluating the impact or visibility of such sites would be to adopt the practice of counting "hits", and focusing only on the "most hit" as the "biggest" sites. But we have seen that most such convenient and efficient practices can all too easily introduce enough bias to seriously cast in doubt the research based on them.

Conclusion

Let me leave the speculative future, and return to the present, and close by noting that the number of conclusions, caveats, and potential research problems connected with studying collaboration in scientific research is enormous. As pleasant and rewarding as it is to solve problems, it is nonetheless even more exciting to realize that there are still more problems to be solved about collaboration, and that there are more problems than there are researchers working on them, which is a good thing for us and the future of our field.

*

Support for this presentation and for travel to this conference was provided by the Office of the Dean of Faculty, Williams College, Williamstown, MA, USA.

References

BEAVER, D. DEB., R. ROSEN, "Studies in Scientific Collaboration, Parts I-III" *Scientometrics*, 1, (1978) 65–84; 1, (1979) 133–149; 1, (1979) 231–245.

BEAVER, D. DEB., Teamwork: A step beyond collaboration, *George Sarton Centennial, Communication and Cognition*, Ghent, Belgium, (1984) pp. 449–452.

KATZ, J. S. B. R. MARTIN, What is research collaboration? *Research Policy*, 26 (1997) 1–18.

KNAPP, R. H., H. B. GOODRICH, *Origins of American Scientists*, University of Chicago Press, Chicago, 1952.

PRICE, D. J. DES., D. DEB. BEAVER, Collaboration in an invisible college, *American Psychologist*, 21 (1966) 1011–1018.

Received July 17, 2001.

Address for correspondence:
DONALD DEB. BEAVER
Department of History of Science, Williams College
Bronfman Science Center, 18 Hoxsey St.
Williamstown, MA 01267, USA
E-mail: dbeaver@williams.edu