

What is co-authorship?

Branco Ponomariov¹ · Craig Boardman²

Received: 4 June 2016 / Published online: 17 September 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract Science and technology policy academics and evaluators use co-authorship as a proxy for research collaboration despite knowing better. Anecdotally we understand that an individual might be listed as an author on a particular publication for numerous reasons other than research collaboration. Yet because of the accessibility and other advantages of bibliometric data, co-authorship is continuously used as a proxy for research collaboration. In this study, a national (US) sample of academic researchers was asked about their relationships with their closest research collaborators—some with whom respondents reported having co-authored and some with whom respondents reported not co-authoring. The results suggest there are numerous dimensions of co-authorship, the most influential of which is informal and relational and with little (directly) to do with intellectual and/or other resource contributions. Implications for theory and practice are discussed. Generally we advise academics and evaluators interested in tracking co-authorship as a proxy for collaboration to collect additional data beyond those available from popular bibliometric resources because such information means better-informed modeling and better-informed policy and management decision making.

Keywords Co-authorship · Research collaboration · Bibliometrics

✉ Branco Ponomariov
branco.ponomariov@utsa.edu
Craig Boardman
pboardma@asu.edu

¹ Department of Public Administration, University of Texas at San Antonio, San Antonio, TX, USA

² Center for Organization Research & Design, Phoenix, AZ, USA

Introduction

For decades now science and technology policy academics and evaluators have been interested in research collaboration. The reasons are varied, but two are dominant. First is a formalized shift in the policy-for-science paradigm from funding individual investigators to funding groups, because presumably the more experts are working together on a particular problem, the better the chances for effectiveness, innovativeness, and/or productivity (Wuchty et al. 2007). Accordingly today many public research investments are made in organized research units with different types of expertise from different economic sectors and/or from different disciplines (Block and Keller 2009). Another (related) reason for academics' and evaluators' interest in research collaboration is that public research investments have become more explicitly problem-focused due to enhanced public and thusly political demand for socially- and/or economically-impactful returns from publicly-funded research (Guston 2000).

A facet of the interest in research collaboration is of course co-authorship. Co-authorship has been operationalized as a proxy for research collaboration since the early 1980s (e.g., Subramanyam 1983). Not just because bibliometric data are widely available, but also because co-authorship often is indeed an output of research collaboration (Melin and Preston 1996). Co-authorship exhibits acceptable face and content validity as a measure of research collaboration—when two or more individuals are listed as co-authors on the same publication, it is quite plausible that these people must have collaborated in some way (Laudel 2002); conversely, it is also implicitly assumed that all scientists who collaborate become co-authors (Beaver and Rosen 1978; Gordon 1980).

Research collaboration seems to be especially notable when co-authors are separated by boundaries, e.g., disciplinary, economic, institutional, generational, gender, national, ethnic, academic, rank, etc. (e.g., Narin and Whitlow 1991; Qiu 1992; Van Raan 1998), because crossing boundaries seems likely to entail the productive integration of different theories, concepts, techniques, and/or data (Porter et al. 2006). In turn, such factors explain why research collaboration is deemed to be of such importance.

Accordingly, substantial body of research modeling research collaboration and using co-authorship as the primary output (or outcome, when co-authorship indeed represents research collaboration) exists. There are numerous studies using bibliometrics to predict the occurrence and intensity of research collaboration with personal/professional and institutional characteristics that are readily gleaned from publications themselves (e.g., Jeong et al. 2011). A large proportion of these studies explain research collaboration in terms of either the resource-based view or more specifically in terms of costs and benefits (Birnholtz 2005; Bozeman and Corley 2004; Melin 2000; Sonnenwald 2007; Traoré and Landry 1997). The literature using co-authored publication as a proxy for research collaboration demonstrates general reliability at the level of individual of analysis (Yan and Guns 2014).

But the current study is not a study of research collaboration. Though we view past bibliometric studies as theoretically informative and also as valuable to policy and management decision making, we suggest here that co-authorship is a social phenomenon worthy of study in-and-of-itself. Though co-authorship may validly represent research collaboration in some instances, co-authorship may have numerous other meanings besides collaboration. This means that co-authorship has yet to be modeled using the appropriate theoretical approach. In this study we seek to remedy this situation (at least in part) by taking theoretical guidance from the law and medical literatures on authorship ascription

rather than from theories of collaborative research (though all three literatures demonstrate some conceptual overlap).

We acknowledge that we are not the first to suggest that bibliometric data are not singly a proxy for research collaboration. In many of the studies of research collaboration that use bibliometric data, as well as in collaboration studies using other types of data, you can find as much, at least implicitly. Typically these studies caveat, usually towards the end of the article, that co-authorship does not necessarily entail research collaboration and/or that many research collaborations produce no publications whatsoever, much less co-authored ones.

Here we hope to begin to move beyond the typical caveats, which in our view requires two steps. The first is understanding what predicts and explains co-authored publications, including but not limited to research collaboration—i.e. to heed Katz and Martin's (1997) call to develop understanding of co-authorship as just a partially valid operationalization of collaboration (Lundberg et al. 2006) but also a valid representation of other phenomena besides collaboration. The second step is to incorporate what we learn about co-authorship as a social phenomenon into future studies of research collaboration. In the current study we take on the first step but are limited to making recommendations for the second.

Thus the paper attempts to unpack co-authorship conceptually. For this task we use data from a national survey of scientists in US universities asking about their closest research collaborators. This is not a bibliometric study but instead relies on survey data eliciting responses about respondents' closest recent collaborations, including responses about co-authorship. The findings suggest that only slightly more than half of collaborations result in co-authorship and, moreover, when this occurs it's probably more a function of relational factors than of intellectual contributions, which is of theoretical importance. A finding important for policy and management decision making is that institutional factors aren't particularly influential.

In the next section we review extant research and theory to inform how to model co-authorship, followed by description of data and methods, and empirical results. The final two sections discuss and synthesize the findings. First we discuss the implications of the findings for theory development and for current policy and institutional efforts to facilitate research collaboration. Last we discuss how to improve upon the current study to help converge theory development and practice regarding co-authorship, research collaboration, and science and technology policy decision making and implementation.

Framing co-authorship

Most study using co-authorship data does so as a proxy for research collaboration (Adams et al. 2003, 2005; Bordons and Gomez 2000; Yoshikane and Kageura 2004; Ponomariov and Boardman 2010). Studies using non-bibliometric operationalizations of collaboration have also proliferated, e.g., based on self-reports from surveys and/or interviews (Boardman and Bozeman 2006; Boardman and Corley 2008; Boardman and Ponomariov 2009; Bozeman and Corley 2004; Kreiner and Schultz 1993; Laudel 2002; Ponomariov and Boardman 2008).

Both of these literatures are relevant to unpacking co-authorship conceptually, albeit with important qualifications. First, studies using bibliometric data either model co-authorship as an outcome of some sort of treatment, such as a new policy and/or institutional affiliation, (e.g., Ponomariov and Boardman 2010) or these studies use it to describe

differences in publication patterns across groups, for instances across disciplines (e.g., Batista et al. 2006). Which means research collaboration studies don't model co-authorship as a discrete phenomenon—our goal in this study. Similarly, studies of research collaboration using self-reports rather than bibliometric data (e.g., Bozeman and Corley 2004) emphasize the relational aspects of collaboration but not the motives and processes of co-authorship per se, which means the theory and modeling from these studies are as directly useful here as are bibliometric studies.

What's needed is a framework specifically for the allocation of authorship credit, what we'll call *authorship ascription* from here on.¹ Though we partially inform our modeling of co-authorship with the research collaboration literature, we are as much influenced by the law and medical literatures on authorship as an indicator of intellectual property and intellectual contribution, respectively. Though these literatures for the most part aren't empirical but instead are predominantly normative and therefore discursive, and though they're focused as much on patents as on scientific publications, their thinking and commentary on authorship ascription as a discrete phenomenon rather than as a proxy for something else (like research collaboration) is directly useful to the current study.

The law and medical literatures on authorship

Biagioli (2003) suggests that authorship is a reward and that it is not per se caused by the intellectual contribution the author makes, meaning that authorship ascription can be a function of things other than intellectual contribution. This argument is consistent with prevailing normative views of scientific activity, i.e. the view that it is the scientific method, not the scientists, that causes scientific results—else, replicability is at stake.² In turn several questions arise: What exactly is authorship ascription a reward for? And when is authorship ascription not a reward but something else?

Biagioli and Galison (2003) conceptualize authorship ascription as a reward for a contribution of capital of some sort (Beaver 2001; Melin 2000; Wray 2002), including financial capital (e.g., a grant award, in-kind contributions of infrastructure and/or technology), human capital (e.g., an intellectual contribution, graduate students), and/or social capital (e.g., being the central node in a new collaboration). Their reasoning is part theory- and research-driven and part anecdotal and intuitive. On one hand there's the resource-based view of research collaboration, which addresses different types of capital contribution as good reason to collaborate (but not necessarily good reason to ascribe authorship). On the other hand we ourselves have been included as authors on papers for making capital contributions.

This reasoning is quite similar to the discourse on authorship ascription in the medical journals like *Journal of the American Medical Association* (Drenth 1998; Flanagin et al. 2002; Lundberg and Glass 1996; Rennie et al. 1997; Riesenbergs and Lundberg 1990; Shapiro et al. 1994). But it comments (generally) that authorship ascription specify the sort

¹ For the remainder of the article we use the term *ascription*. We avoid using the term *attribution* so as not to confuse our efforts here with the long- and well-established authorship attribution literature, which focuses on inferring the unobservable characteristics of an author from the content of the documents written by that author.

² It would be worthwhile to consider a more in-depth investigation on the possible qualitative differences between single- and multi- authored papers. Anecdotally, it seems that highly influential, paradigm-shifting theoretical papers tend to be single-authored, while papers involving empirical, "normal science" work are more-likely, and better if, co-authored in comparison. In the qualitatively different case of individual theoretical breakthroughs, it appears to easier to justify to apply the authorship-as-cause view.

of capital contributed. Intellectual oversight, such as that by a mentor, should be ascribed when the paper is the product of a particular individual's protégés and/or other individuals who required guidance and advice during the research. These contributors should be designated as *guarantors* not authors, and those who make direct intellectual contribution to the research described in the manuscript should be called *contributors* (and their specific contribution should be specified in the credits at the beginning of the paper) (e.g., Rennie et al. 1997).

The International Committee of Medical Journal Editors is much more systematic. They argue that if but one of the follow criteria are not met by an individual, then that individual should be acknowledged as a contributor (e.g., in a footnote) but not ascribed authorship. These criteria are as follows (verbatim):

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
- Drafting the work or revising it critically for important intellectual content; AND
- Final approval of the version to be published; AND
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.³

Because the extent to which the above conditions are met by any individual authors is practically impossible to verify, and because as far as we know no systematic attempts are made to audit the appropriateness of authorship ascription, these criteria are no more than a normative set of recommendations. The obvious verification and measurement validity and reliability issues with these criteria notwithstanding, it is notable that social and relational reasons unrelated to capital contributions for ascribing authorship are omitted. Which implies that relations alone are no reason to ascribe authorship. Yet, we know from the research collaboration literature that is survey—not bibliometrics-based that relationships matter and can sometimes lead to authorship ascription, including unwarranted ascription per the above-listed criteria (Bozeman and Youtie 2015).

Just as important, the absence of actions not directly related to producing the co-authored research from the above list, fundamentally challenges the assumption of congruence between collaboration and co-authorship, because *a shared research goal is not a necessary presupposition for collaboration* (Laudel 2002: 5); actors may collaborate according to their interests, do not necessarily share goals, and the collaboration is defined by the activities involved—and the range of these activities (some of which may be rewarded by co-authorship)—is much broader than the short check list provided above.

From a theoretical point of view, studying the relational determinants of co-authorship is important because the resource-based and institutional views, though intuitively appealing, by definition are incomplete. Both views assume that co-authorship is of a somewhat discrete quality (rather than a function of ongoing relations)—meaning that co-authorship happens in order to satisfy identifiable needs and not as a function of past and contemporaneous experiences (including but not limited to past and current relations). Accordingly from these views scientists and engineers merely scan their environments for possible collaborators (and co-authors) based on their particular capital needs and/or based on the particular institutional influences they're subjected to at any given time. These views implicitly deny that non-resource relations can help to predict and explain authorship ascription.

³ Viewed on 01 May 2016; <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>.

The research collaboration literature

The research collaboration literature using bibliometric data [sometimes complemented by other methods (Katz and Martin 1997)], similarly adopts an institutional and/or resource-based view, to the exclusion of relational predictors. These studies also make two assumptions: (1) that co-authorship indeed signifies actual research collaboration and (2) that all scientists who collaborate become co-authors. While the first assumption is often violated (Harsanyi 1993), it is also relatively easy to address statistically (Laudel 2002). The second assumption is much more problematic, because co-authored publication depicts only a fraction of collaborative activities (Melin and Persson 1996).

The resource-based and institutional explanations of bibliometric studies of research collaboration have reasonable face validity, but, similar to the law and medical explanations (see “[The law and medical literatures on authorship](#)” section) are also likely insufficient. First, approaching collaborations strictly from an incentives/assets/structure perspective pays insufficient attention to the possibility that at any given time, there is a greater number of possible collaborations than actually occur. Yet scientists collaborate with a limited number of colleagues—arguably with far less than they could in principle, based on the likelihood that from the resource point of view the universe of suitable collaborators is larger than the actual set of collaborators they interact with. If that is the case, then it suggests that the relationships extend beyond discrete projects, do not necessarily involve shared goals, and the explanation of co-authorship may also have something to do with the relational characteristics of these general collaborative relationships rather than any specific incentives as perceived on a case by case basis.

Thus, research collaboration is a multi-dimensional process, of which co-authorship is only one potential dimension. If collaboration is such a fluid and hard to demarcate process, then resource-based and institutional explanations are perhaps less important, simply because an on-going collaborative relationship (much like a marriage) is sustained not on the basis of discrete exchanges (and regular dissolution and re-engagement), but rather on mutual and multidimensional social commitment that is neither easily established nor readily terminated.

We see some of this in the research collaboration literature not using bibliometric data. Particularly in the studies using surveys of academic researchers’ values and motives (see Kingsley et al. 1996 for an overview). Yet these studies emphasize as dependent variables a slew of attitudes and perceptions that don’t speak directly to outputs like publications (e.g., Bozeman and Corley 2004).

Accordingly, the main focus of our study is on examining the likelihood of co-authorship, based on a data set that captures the broader multi-dimensionality of research collaborations. Specifically, doing so requires (1) accepting that collaboration is a fluid and multi-dimensional process, and (2) that this multidimensionality (i.e. the various aspects of the collaborative relationships) has direct relevance to the actual knowledge production process (e.g., producing publications). While scientists do pursue knowledge for its own sake and in the process only collaborate with others, the knowledge produced is not ultimately legitimate unless it survives some form of the peer review process and appears in print. Hence, the main question of this paper is not open-ended, but identifies specific, theory-driven explanation of what properties of the collaborative relationships increase the likelihood of a co-authored publication.

The implicit general hypothesis underlying this question is that scientists do not co-publish with each other for specific reasons, e.g., like resources, but rather that the process

requires a more complete explanation, featuring both resource-based and relational factors. The sections below propose that the scientific and technical human capital framework—in our view which has been insufficiently operationalized to date—provides a robust template to integrate both explanations empirically. Because this approach helps to reconcile many of the validity issues associated with studying collaboration through co-authorships.

First, it recognizes that collaboration does not necessarily always lead to co-authored papers, it recognizes that it can also lead to other outputs, or nothing tangible at all. Conversely, the approach employed here also limits another common problem in bibliometric studies, namely the fact that not all co-authorships are the result of actual collaboration: the data analyzed in this study by definition only incorporate co-authored publications that have resulted within the context of a collaborative relationship. As a result, the proposed study focuses on explaining the likelihood of co-authored publication within an existing collaboration's context, thus capturing collaborations—and aspects of collaborations—that cannot be measured through co-authorships. This approach also provides reassurance that the co-authored publications are the result of actual collaborative relationship. Finally, it provides an illustration of the feasibility to study productivity patterns without sacrificing important complementary information regarding the nature of the collaboration.

Hypotheses

We emphasize both relational and resource-based predictors of co-authorship. However, we characterize both as ultimately relational. Resources and relations are not separable: there typically cannot be the former without the latter of some sort (Corley and Bozeman 2004; Boardman 2009; Ponomariov and Boardman 2010). For relational collaborations, the relation itself could be a hidden contribution of human capital-augmented labor between collaborators, who in the process of the relation could receive useful contributions that they cannot necessarily fully reproduce on their own. For resource-based relations, we distinguish between the formal and informal while acknowledging that formal relations based on the provision or exchange of resources can also have informal resource-based elements as well as relational elements that have little to do with resources per se (Bozeman et al. 2001). In other words, and as usual, no social phenomenon is ever observed in its pure form, although the categories of relations can be conceptualized as discrete.

Formal resource-based relations

Resources (or capital) broadly defined are what most of the above-discussed literatures emphasize. Even the tangentially-related policy analysis literature (tracking co-authorship longitudinally) addresses resources in that many of the policies assessed for ex post bibliometric impact entail the allocation of financial and other resources (to the subjects being examined for co-authorship). The first formal resource-based relation we propose as a determinant of co-authorship is mentoring relations:

H1 In collaborations wherein one of the collaborators is formally a mentor of another of the collaborators (e.g., dissertation adviser, dissertation committee member, or other formal mentorship roles), co-authorship is more likely.⁴

⁴ We don't get more specific than this due to response bias and measurement reliability (Bozeman et al. 2012). First, collaborators can have very different views of their own and others' contributions to the same

Formal resource-based relations as a determinant of co-authorship is straightforward and has been elucidated in the numerous literatures on research collaboration discussed above. Given our emphasis of alternate explanations of co-authorship, we allocate more space discussing the informal and relational predictors rather than reiterating the resource-based view (which emphasizes formal resources).

Informal resource-based relations

Research collaborations also entail resource-based relations that aren't formalized, which makes the predominance of research collaborations in academia informal and difficult to detect empirically, and many aren't well-studied (Hagedoorn et al. 2000). What's needed is a way to identify when collaborations entail resource-based relations in the absence of formalization.

One approach is to emphasize boundaries, e.g., institutional, economic, and disciplinary boundaries. The basic idea is the more boundaries are spanned, the more productive and more diverse is the collaboration (Boardman 2009; Bozeman and Corley 2004; Ponomariov and Boardman 2010). Assuming researchers from firms bring different resources to collaboration than bring researchers from academia; researchers trained in biometrics bring different resources than bring researchers trained in psychology; and so forth. However, spanning boundaries also entails a cost, sometimes prohibitive, making interdisciplinary or inter-institutional collaborations costly, difficult or sometimes even plainly impossible. In collaborations that have actually occurred (as in those studied in the present paper), willingness to bear this cost suggests that the payoff of the collaborative relationship, or the resources contributed, outweighs this cost (e.g., communication or coordination difficulties). The key assumption remains that authorship is ascribed when resources are contributed.

H2 When collaborators are separated by boundaries (e.g., different universities, different economic sectors, different disciplines, different generations), co-authorship is more likely.

Of course, identifying boundaries as a proxy for informal resource relations is blunt. For example researchers from different fields, sectors, and/or institutions could be resource/capital substitutes not complements yet no less included in the collaboration for non-resource (i.e. purely relational) reasons (see "[Non-resource relations](#)" section). And/or there can be intra-institutional, -disciplinary, -sector relations that are informal yet resource-based. In other words the boundaries approach to measuring informal resource contributions has the potential for high rates of both false positives and false negatives.

An alternate approach is to consider informal human capital contributions that occur in professional relationships (Corley et al. 2006; Laudel 2001; Mullins 1973).

H3 When a researcher views his or her collaborator as someone who can assess and help to improve his or her research (e.g., by reviewing results), the likelihood of co-authorship goes up.

Footnote 4 continued

project. Related, beyond formalized mentor–mentee relations (e.g., once mentees graduate and become "equal" to their advisers), it's difficult to reliably track formalized resource-based relations amongst collaborators, for instance amongst co-PIs. This is one of the reasons the analysis below also includes separate equations by academic rank, to control for the changing context of individual careers and collaboration roles, especially clear with regards to relationship with mentors.

Non-resource relations

Many of the predictors we characterize as non-resource relations of course may coincide with resource-based relations. Many potential predictors are not readily validated as “resource-based” (versus those just addressed⁵) but are valid relational variables. Meaning that they can help to characterize relationships between or amongst collaborators and thusly to distinguish one research collaboration from the next.

H4 The more frequent the communication between collaborators (e.g., same institutional affiliations, same conferences, self-reports of communication), the more likely there will be a co-authored publication.

H5 The “closer” the collaborators (e.g., length of relationship, professional trust and respect, friendship outside work) have known one another, the more likely will there be a co-authored publication.

A number of prior studies (Hagstrom 1965; Katz and Martin 1997; Price and Beaver 1966) suggest that most research collaborations start informally, as a result of informal conversations at conferences and so forth, which may “stimulate them to think about unsolved problems in their field, about possible research projects, about the interpretation of older data and the like” (Laudel 2001: 8). Our main assumption here is that non-resource relations increase the likelihood of co-authored publication, perhaps via formal resource-based relations, and also likely (at the least) via informal resource-based and non-resource relations.

Data description and key variables

Survey and co-authorship data

The data reported in this paper comes from the first stage of a large 3-year study of social and collaborative networks of scientists and engineers.⁶ The multi-year effort includes a large national two-stage survey of US academic scientists and engineers in six fields. The data reported here are drawn from the first-stage survey which was completed in March 2007. This survey also captures the structure of collaborative and advice networks by investigating the connections between the collaborators named by the respondents.

The survey included three major categories of questions. First, the most extensive of these questions was a series of name generator and name expander questions based on research methods typical to sociological studies of social networks. The name generator questions were used to identify key collaborators or advisors in several key categories, including formal as well as research advice networks. Because an important focus of the study is on the social aspects of the academic enterprise, two name generators were used to identify “collaborative networks”: individuals inside and outside of the university with

⁵ For example being a graduate student or adviser of a graduate student constitutes verifiable formalized human capital, social capital, and capital resource relations. Inter-organizational research partnerships are entered predominantly for resource-based reasons. And an individual’s view that a collaborator contributes intellectually as a guarantor is also valid. See “[Formal resource-based relations](#)” and “[Informal resource-based relations](#)” sections.

⁶ “Women in Science and Engineering: Network Access, Participation, and Career Outcomes,” a project funded by the National Science Foundation (Grant# REC-0529642).

whom the respondent works on research. While the name generators are useful for identifying collaborators, it was also important to understand characteristics of these individuals, and the characteristics of their relationships with the respondent. To do this, a series of “name expander” questions were used to capture the nature of the collaboration (nature of research product), details about the level of relationship and origin of acquaintance, closeness of research expertise, communication frequency, grant activity, and general demographics. The survey was implemented online using Sawtooth Software[®]. Individuals were alerted to the survey via personal email and provided with a unique user id and password (and directed to the website). Three reminders were sent, with a combination of email and postcard reminders.

The survey sample of 3677 was randomly drawn from the population of academic scientists and engineers in six disciplines in Carnegie-designated Research I universities (151 universities at the time of the survey). The sample was stratified by gender, rank (i.e. assistant, associate or full professor) and discipline.⁷ Overall, 1764 survey were returned for a 50.1 % response rate and a usable response rate of 47 %.⁸ Responses were fairly evenly distributed across the six fields, gender (48 % women) and rank (26 % assistant professor, 28 % associate professor, and 46 % full professor). Emeritus, research scientists and any scientists that reported not being in tenured or tenure-track positions were removed for this analysis, resulting in a final total sample size of 1581. Descriptive statistics for the respondents used in the present paper are presented below in Table 1. On average, respondents named 4.6 close collaborators. Respondents were more likely to nominate close collaborators outside of the university than inside: on average, respondents had 2.7 close collaborators inside of their home university and about 3 close collaborators outside of their university; the difference is statistically significant ($P = 0.0001$).

The survey asked all scientists to respond to a close-ended question of how many articles they (personally) have published in the last 2 years. The options given were “0”, “1–2”, “3–4”, “5–6”, “7–9”, “10–14”, and “>15”, i.e. the number of publications variable is ordinal. Since the number of categories is large, the data was recoded in three productivity groups (low, medium and high) based on the tertile distribution.⁹ Accordingly, the observations in the first tertile fall into the “4 or less” publications group, the middle tertile results in the “5–6” publications group, and the third tertile encompasses the “more than 7” publications group. See Table 1 for the shares of respondents falling in each group. Approximately half of all respondents fall in the “low” productivity group, 20 % in the middle, and 30 % in the “high” productivity group.

Unit of analysis: the collaboration, not the individual

Although respondent level control variables are used in the modelling of collaborations, the individuals themselves are not the unit of analysis in this study. The units are the closest

⁷ Disciplines were selected based on the level of female representation in order to allow for comparison across male-dominated versus gender balanced fields of study. The fields were chemistry, biological sciences, earth and atmospheric sciences, electrical engineering, computer science, and physics.

⁸ Data were cleaned for incomplete responses. In the cleaning, non response due to bad addresses were also removed for the calculation of response rate. For example, 136 of the emails were “bounced back” due to a bad email address and 19 were “returned to sender” by the recipient universities email server. Follow-up calls were made but respondents could not be located in these cases.

⁹ The division based on tertile values is arbitrary and chosen to limit the number of separate equations run for the separate productivity groups. Also, since for theoretical purposes a general delineation of levels of productivity suffices, more granular division did not seem practical or necessary.

Table 1 Descriptive statistics for individual respondents

Variable	Observations	Mean	SD	Min	Max
Rank					
Assistant professor	1581	26.1 %	0.44	0	1
Associate professor	1581	27.5 %	0.45	0	1
Full professor	1581	46.4 %	0.50	0	1
Total	1572	100 %			
Publications					
Number of journal articles published in the last 2 years: 4 or less	1572	48.4 %	0.50	0	1
Number of journal articles published in the last 2 years: 5–6	1572	19.1 %	0.39	0	1
Number of journal articles published in the last 2 years: 7 or more	1572	32.5 %	0.47	0	1
Total	1572	100 %			
Discipline					
Biology	1581	17.5 %	0.38	0	1
Chemistry	1581	18.0 %	0.38	0	1
Computer science	1581	16.4 %	0.37	0	1
Earth/atmospheric sciences	1581	18.0 %	0.38	0	1
Electrical engineering	1581	13.4 %	0.34	0	1
Physics	1581	16.8 %	0.37	0	1
Total	1581	100 %			
Number of closest collaborators at respondent’s university	1321	2.7	1.41	0	5
Number of closest collaborators outside of respondent’s university	1285	2.9	1.47	0	5

recent (last 2 years) collaborations provided by the respondent. As a part of the survey, respondents were asked to name their *closest research collaborators* (both in their own university as well as outside of the university). They could name up to 5 collaborators in each of these two groups. The exact wording of the questions regarding the nomination of the collaborators is presented in the “[Appendix](#)”. Overall, respondents were given minimum guidance in how to define “collaboration”, although they were prompted to think about tangible work relationships resulting in a certain output. Precisely operationalizing collaboration is very challenging (e.g., [Laudel 2002](#)). Instead, a plausible alternative is to simply leave the scientists themselves to identify their “closest research collaborators”, and then ask follow-up questions about these relationships. Obviously, this approach by design prioritizes “true” or “deep” collaborations, while ensuring that casual, unimportant, or opportunistic collaborators are less likely to be nominated.

The 1581 respondents used in this analysis named a total of 7272 collaborators. After naming the collaborators, respondents were asked to provide detailed information about their collaborative relationship with each particular person they named. After listwise elimination of observations with missing values, the collaboration data set is comprised of 5621 observations each describing a specific collaborator and the collaborative relationship. Since the fundamental premise of the paper is that characteristics of the collaborative relationship affect the outputs of the relationship this data format is most effective in

testing hypotheses of such relationships. Table 2 provides the descriptive statistics for the collaboration unit of analysis.¹⁰

It is notable that only about half of the collaborations had resulted in co-authored journal article—an estimate consistent with prior study assessing that about half of collaborations remain invisible in formal publication channels (Laudel 2002).

Key variables

The dependent variable is a binary based on self-reports of whether or not the collaboration resulted in a co-authored publication. We estimate a general model for the entire sample, and then a series of separate models for sub-groups defined by respondent's rank (assistant, associate, or full professor), as well as by productivity (low, medium, and high), in order to guard against the possibility that the processes of collaboration and co-authored publication are qualitatively different at different levels of seniority and experience, as well as overall productivity and capacity.

The variables for hypothesis 1 on formal resource-based relations are based on questions regarding the status of the collaborator, broadly defined, relative to the status of the respondent. For example, positive status differential might increase the likelihood of publication. Accordingly, operationalizations of H1 include questions asking if the collaborator was on respondent's dissertation committee, if he or she is a PhD student of the respondent.

The variables for informal resource-based relations (hypotheses 2–3) are based on questions operationalizing whether respondents and their collaborators inhabit the same institutional or generational spaces. Accordingly, they include questions about whether the collaborator is senior or junior to the respondent, if they are from outside of their university, and the extent to which respondents understand the collaborator's area of specialization.

Hypothesis 3 uses survey item asking about guarantorship, e.g., different forms of informal intellectual or other contributions. It includes questions on whether the collaborator have explicitly reviewed and recommended improvements to respondent's research, and whether the collaborator has introduced the respondent to other collaborators, or nominated him for awards.

¹⁰ Conceptualizing the unit of analysis in this way combines the concept of a "collaboration" broadly defined with a more explicit look at the characteristics of the "relationship" that underpins this collaboration. In other words, such conceptualization allows a broader operationalization of the concept of scientists' collaborations understood as a complex phenomenon anchored around a relation—i.e. an interaction between two agents that is based on specific mechanisms, and actions, and is also contingent on the properties of the agents forming the relationship. This allows simultaneously broadening the lenses through which to observe the relationships underlying collaborations (i.e. as encompassing more than co-authorships), yet keeping it focused enough so that the phenomenon of collaboration remains articulated and measurable through the incidence of specific relationship mechanisms and properties. Respondents were asked to nominate individuals whom they consider to be their "closest collaborators." No operational definition of a "collaborator" was given to respondents, in order to elicit nominations based on respondent's perceptions of what kinds of relationships constitute "close collaboration." Once a respondent had identified a collaboration, it was possible to gather information about specific properties of such relationships, which in turn allows examining the extent to which the selected predictors (see "Hypotheses" section) co-vary with co-authorship. Similarly, given that the unit of analysis is each individual collaborative relationship (rather than the individual), the models below, as specified in "Hypotheses" section examine what properties of the relationship would affect the likelihood that it will yield a publication within the last 2 years. Thus the single dependent variable for this study and in the models below is a binary variable coded 1 if the collaboration has yielded a publication within the last 2 years, zero otherwise.

Table 2 Descriptive statistics of dyadic collaborations

Variable	Observations	Yes (%)	No (%)	Total (%)
Formal resource based relations				
This person was on my dissertation committee	6332	3.0	97.0	100
This person was/is a Ph.D. student of mine	6332	4.3	95.7	100
Collaborated on a journal article ^a	5763	53.4	46.6	100
Collaborated on research grant proposal	5764	56.5	43.50	100
Collaborated on conference paper	5763	50.5	49.5	100
Collaborated on a patent	5763	3.9	96.1	100
This person has nominated me for an award or as an invited speaker	6326	16.3	83.7	100
Informal resource-based relations				
Collaborator is outside of the university	7253	51.5	48.5	100
I have a detailed understanding of this person’s area of specialization* ^b	6289	52.1	47.9	100
I have a working understanding of this person’s area of specialization	6289	44.3	55.7	100
I have little to no understanding of this person’s area of specialization	6289	3.5	96.5	100
This person has reviewed papers on which they are not a co-author prior to submission	6326	26.2	73.8	100
Collaborator is senior to me	6334	38.9	61.1	100
Collaborator is junior to me	6334	25.2	74.8	100
I met this person for the first time at a conference	6332	17.7	82.3	100
This person has introduced me to potential collaborators	6326	32.7	67.3	100
Non-resource relations				
Collaborator is a close friend	6334	26.1	73.9	100
Frequency of communication: about daily*	6232	12.9	87.1	100
Frequency of communication: about weekly	6232	30.2	69.8	100
Frequency of communication: about monthly	6232	32.3	67.7	100
Frequency of communication: less often than monthly	6232	24.5	75.5	100
Length of the relationship: <3 years*	6332	15.2	84.8	100
Length of the relationship: 3–6 years	6332	31.9	68.1	100
Length of the relationship: >6 years	6332	52.9	47.1	100
We were Ph.D. students together with this person	6332	3.6	96.4	100
Discipline controls				
Biology*	7253	16.0	84.0	100
Chemistry	7253	15.8	84.2	100
Computer science	7253	16.3	83.7	100
Earth/atmospheric sciences	7253	21.8	78.2	100
Electrical engineering	7253	13.5	86.5	100
Physics	7253	16.6	83.4	100

All variables are binary. Standard deviations are not reported as they have no substantive meaning

^a Dependent variable

^b All variables marked with an asterisk represent the reference groups in the logistic regression models in the following section

Table 3 Logit models: a collaboration has resulted in a journal publication over the last 2 years (dependent variable)

	(1)	(2) ^a	(3) ^b	(4) ^c	(5) ^d	(6) ^e	(7) ^f
Formal resource-based relations							
This person was on my dissertation committee	0.103 (0.185)	0.930*** (0.288)	-1.121*** (0.372)	-0.319 (0.463)	0.075 (0.263)	0.015 (0.317)	0.592 (0.577)
This person was/is a Ph.D. student of mine	0.546*** (0.173)	1.189* (0.664)	0.050 (0.371)	0.631*** (0.208)	0.437* (0.245)	0.892*** (0.352)	0.435 (0.375)
Collaborated on research grant proposal	-0.336*** (0.062)	-0.337*** (0.125)	-0.392*** (0.119)	-0.261*** (0.093)	-0.369*** (0.096)	-0.332*** (0.111)	-0.433*** (0.139)
Collaborated on a conference paper	-0.087 (0.059)	-0.300** (0.120)	-0.122 (0.112)	0.009 (0.089)	-0.082 (0.092)	-0.151 (0.105)	0.036 (0.128)
Collaborated on a patent	0.536*** (0.157)	0.412 (0.319)	0.817*** (0.290)	0.440* (0.241)	-0.118 (0.280)	0.685*** (0.278)	0.875*** (0.311)
This person has nominated me for an award or as an invited speaker	0.173** (0.082)	-0.150 (0.179)	0.040 (0.153)	0.306** (0.123)	0.062 (0.134)	0.020 (0.145)	0.241 (0.170)
Informal resource-based relations							
Collaborator is outside of the university	0.670*** (0.075)	1.075*** (0.158)	0.676*** (0.151)	0.523*** (0.108)	0.706*** (0.119)	0.691*** (0.132)	0.705*** (0.162)
I have a working understanding of this person's area of specialization	-0.354*** (0.063)	-0.165 (0.132)	-0.281** (0.118)	-0.537*** (0.093)	-0.484*** (0.099)	-0.366*** (0.110)	-0.138 (0.137)
I have little to no understanding of this person's area of specialization	-0.747*** (0.180)	-0.810** (0.379)	-0.790** (0.382)	-0.851*** (0.253)	-1.105*** (0.322)	-0.964*** (0.300)	0.033 (0.392)
This person has reviewed papers on which they are not a co-author prior to submission	0.161** (0.070)	0.295** (0.135)	0.220 (0.134)	0.129 (0.111)	0.217** (0.108)	0.078 (0.123)	0.262 (0.161)
Collaborator is senior to me	-0.088 (0.069)	-0.188 (0.141)	0.187 (0.128)	0.077 (0.121)	-0.048 (0.110)	0.008 (0.120)	-0.006 (0.153)
Collaborator is junior to me	0.320*** (0.069)	0.233 (0.141)	0.532*** (0.128)	0.123 (0.121)	0.158 (0.110)	0.531*** (0.120)	0.254 (0.153)

Table 3 continued

	(1)	(2) ^a	(3) ^b	(4) ^c	(5) ^d	(6) ^e	(7) ^f
I met this person for the first time at a conference	(0.080)	(0.252)	(0.166)	(0.104)	(0.128)	(0.141)	(0.169)
	-0.138*	-0.355**	-0.064	-0.057	-0.297**	-0.163	0.212
	(0.078)	(0.164)	(0.150)	(0.116)	(0.123)	(0.137)	(0.177)
This person has introduced me to potential collaborators	0.146**	0.529**	0.052	-0.017	0.201**	0.204*	0.037
	(0.065)	(0.134)	(0.122)	(0.097)	(0.100)	(0.118)	(0.142)
Non-resource relations							
Collaborator is a close friend	0.139*	0.063	0.122	0.247**	0.217*	0.272**	-0.162
	(0.074)	(0.154)	(0.138)	(0.111)	(0.116)	(0.126)	(0.164)
Frequency of communication: about weekly	-0.113	-0.489**	-0.171	0.015	-0.200	-0.390**	0.115
	(0.100)	(0.213)	(0.188)	(0.148)	(0.152)	(0.196)	(0.207)
Frequency of communication: about monthly	-0.341***	-0.738***	-0.575***	-0.048	-0.447***	-0.646***	-0.179
	(0.110)	(0.230)	(0.210)	(0.162)	(0.169)	(0.213)	(0.228)
Frequency of communication: less often than monthly	-0.595***	-0.876***	-0.735***	-0.479***	-0.586***	-0.855***	-0.668**
	(0.124)	(0.256)	(0.246)	(0.180)	(0.191)	(0.234)	(0.262)
Length of the relationship: 3–6 years	1.021***	1.178***	0.764***	0.866***	1.073***	0.900***	0.831***
	(0.095)	(0.153)	(0.193)	(0.178)	(0.153)	(0.161)	(0.220)
Length of the relationship: >6 years	1.134***	0.978***	1.061***	0.783***	1.030***	1.129***	0.980***
	(0.094)	(0.187)	(0.195)	(0.171)	(0.152)	(0.159)	(0.219)
We were Ph.D. students together with this person	0.320*	0.541*	-0.020	0.840**	0.254	0.319	0.388
	(0.167)	(0.296)	(0.277)	(0.354)	(0.248)	(0.298)	(0.377)
Discipline controls							
Chemistry	0.339***	0.354*	0.304*	0.286*	0.296*	0.570***	-0.501**
	(0.097)	(0.201)	(0.183)	(0.147)	(0.158)	(0.179)	(0.212)
Computer Science	-0.002	0.466**	-0.128	-0.211	0.464***	0.107	-0.994***
	(0.101)	(0.207)	(0.197)	(0.149)	(0.141)	(0.190)	(0.306)
Earth/atmospheric sciences	0.284***	0.430**	0.349*	0.169	0.394***	0.284*	-0.222

Table 3 continued

	(1)	(2) ^a	(3) ^b	(4) ^c	(5) ^d	(6) ^e	(7) ^f
Electrical engineering	(0.093)	(0.188)	(0.181)	(0.139)	(0.144)	(0.164)	(0.226)
	-0.038	0.364	-0.223	-0.098	0.053	-0.043	-0.635**
	(0.112)	(0.238)	(0.196)	(0.177)	(0.180)	(0.196)	(0.257)
Physics	0.299***	0.726***	0.024	0.161	0.367**	0.475***	-0.701***
	(0.096)	(0.187)	(0.192)	(0.146)	(0.162)	(0.173)	(0.210)
Constant	-0.827***	-1.105***	-0.741**	-0.359	-1.194***	-0.478*	0.196
	(0.159)	(0.302)	(0.313)	(0.258)	(0.247)	(0.289)	(0.360)
Observations	5621	1549	1524	2548	2358	1917	1331
Log likelihood	-3512	-888.5	-975.7	-1578	-1429	-1157	-782.2
LR χ^2	742.1	358.9	160.2	255.1	348.5	290.1	130.4
Pseudo R ²	0.096	0.168	0.076	0.075	0.109	0.111	0.077

^a Assistant professors only

^b Associate professors only

^c Full professors only

^d Low productivity respondents only (4 publications or less in the last 2 years)

^e Medium productivity respondents only (5 or 6 publications in the last 2 years)

^f High productivity respondents only (more than 6 publications in the last 2 years)

The variables for non-resource relations (hypotheses 4–5) include survey items assessing the social aspects of the relationship, including the length of time collaborators have known each other, the frequency of communication, and whether the collaboration is based on friendship.

Results

The results support each of the hypotheses, most notably the proposition that authorship ascription is informal and relational than for the proposition that ascription is resource-based (Table 3).

Formal resource-based relations (hypothesis 1)

The results generally provide a weak support for the first hypothesis (that capital contributions increase the likelihood of co-authored publication), with some qualification. If a collaborator was on a scientist's dissertation committee, it is more likely that the relationship has yielded a co-authored publication. However, this dynamic only applies to the subset of assistant professors—i.e. junior scholars who can be expected to still work closely with their advisors from graduate school. Predictably, this relationship not only does not apply in any of the other rank/productivity subsets, but for the case of associate professors it is negative and significant; for example, faculty whose central collaborators remain their advisors mid-career may lack independence or initiative, which may eventually hurt their overall productivity and the productivity of existing collaborations. Together, these results suggest that advisor-student relationships are important at the onset of one's career, however unless the junior scientist adopts an independent path soon after, these relationships become less and less productive in terms of co-authorship over time.

Somewhat different is the effect of the collaborator being a current or former student of the respondent. The findings show somewhat inconsistent evidence that such collaborations are more likely to result in a co-authored publication. This reasoning is warranted since working with a graduate student—current or former—almost by definition implies some co-authored publication activity, which in this case seems to apply only to the low- and medium-productivity scholars, as well as to assistant and full professors.

Having been graduate students together is weakly and positively associated with the likelihood of co-authored publication, a relationship discernible only in the context of assistant and full professors. In the former, perhaps because fellow former graduate colleagues constitute the most readily available pool of collaborators; for the latter, perhaps because it is more likely that graduate student friendship that has been sustained over the entire career continue to be productive.

Informal resource-based relations (hypotheses 2–3)

The findings show mixed support for the second hypothesis (that the more boundaries spanned, the more likely is co-authored publication). Across all models, *collaborations in which the collaborators are in a different university than the respondent's are more productive with regards to publications*. Specifically, such collaborations are 15 percentage points more likely to result in co-authored publication than same-university collaborations. This suggests that scientists are willing and able to pursue potentially productive

collaborators exclusively on merit, rather than to rely on institutional facilitation, and when boundaries are crossed, it is for a “good reason”, outweighing the implicit costs of boundary-spanning. This finding is not surprising however, given that the choices for collaborators outside one’s own department and university is always larger than the options available within the institutional boundaries of the respondent’s department or university. Indeed, collaborators from outside of the university may be more professionally compatible with the respondents than their colleagues—being in a different university does not mean being in a different discipline, or having a different specialization. In fact, the similarities in knowledge and backgrounds may far outweigh the relatively minor inconvenience of collaborating across institutional boundaries, which is supported by the findings regarding the effect of similarity of knowledge and backgrounds on the likelihood of co-authored publication.

In particular, having different backgrounds and training makes collaborations less likely to result in a co-authored publication. The less detailed a scientist’s understanding of his or her collaborators’ area of specialization, the less likely that they will co-publish a paper together, and the lesser the extent of understanding (i.e. “I have little to no understanding...” vs. “I have a working understanding...”), the lesser the likelihood of a co-authored publication, and vice versa—i.e. the easier the mutual understanding, the lower the cost of co-authored publication.¹¹ While the latter universally decreases the likelihood of co-authored publication across all models, the former presents interesting exceptions in the case of assistant professors and highly productive individuals, where in both cases having a “working understanding” of collaborator’s specialization does not negatively affect the likelihood of publication, albeit probably for different reasons. Perhaps assistant professors are simply less discriminating in whom they collaborate with, while highly productive individuals’ competences may also include the ability to collaborate with diverse set of specializations.

As for hypothesis 3 (that when a collaborator is a qualified “assessor” or “guarantor” of the quality of one’s research, the likelihood of co-authorship increases) being a reviewer generally shows a positive effect on the likelihood of co-authored publication, however only for the two subgroups—assistant professors and low-productivity individuals, i.e. precisely lower status groups that can benefit the most from guarantorship. This may also suggest that collaborations in such context tend to be more organic and social than collaborations of advanced and highly productive scientists. Virtually the same effect can be attributed to a collaborator introducing the respondent to potential collaborators, except that the positive effect of this behavior on co-authored publication also extends to medium-productivity collaborators. Last, having a collaborator nominate the respondent for an award or as an invited speaker only has a positive effect on co-authored publication in the case of full professors, a finding perhaps consistent with the finding on “friendship” and suggesting that close social relationships may be the consequence of a long collaborative career, rather than a precondition.

Finally, one of the central, and unambiguously finite, resources that collaborators share is simply time. Any aspect of collaboration involves allocations of time, and some time allocations necessarily reduce the amount of time available to devote to research leading to co-authored publication. In particular, having collaborated on a grant proposal over the last 2 years is negatively related to likelihood of co-authored publication over the same time period across all models, at high levels of significance. Proposal writing requires

¹¹ Per a (post-estimation) Wald test, the differences between the coefficients is statistically significant ($P = 0.024$).

substantial time investment which can only come from other time-intensive activities, such as the time needed to research, experiment, and write up the results. Thus although grant-writing activity might eventually allow for greater productivity in the future, in the short term it appears to reduce the productivity of collaborations.

Non-resource relations (hypotheses 4–5)

The findings show mixed support for the fourth and fifth hypotheses (that the frequency of contact and personal “closeness” increase the likelihood of co-authored publication, respectively). As hypothesized in the introduction, collaborative relationships characterized with intensive and regular information exchanges should be expected to positively affect the probability that such a collaboration will culminate in a published paper. The models tested in this study universally support such reasoning, with an interesting caveat. In the general model for the entire population (Model 1), the relationship between frequency and likelihood of having co-authored a journal publication is positive: the higher the frequency, the greater the likelihood. Specifically, using daily frequency of communication as a reference group,¹² communicating about weekly results in no discernible change in the likelihood of co-authored publication, while communicating about monthly or less often reduces the probability of co-authored publication by 7 and 11 percentage points respectively, keeping all other variables constant at 0. The result is plausible, insofar communication that is less frequent than at least a monthly contact may signify either that the collaborators are too busy with other activities, and hence the productivity of the collaboration declines, or that the collaboration itself is of lower priority.

This relationship holds across all models for all subgroups of interest, with two notable exceptions: full professors (Model 4) and high-productivity respondents (Model 7). In both cases, reduced frequency of communication does not negatively affect the likelihood that the collaboration has resulted in co-authored publication, unless it less often than monthly thereby suggesting a somewhat higher level of “automaticity” or “stability” of the collaboration. Conversely, two contexts in which collaborations appear to be particularly sensitive to frequency are the collaborations of assistant professors (Model 2) and the collaborations of medium productivity scholars (Model 6). In both cases reduced frequency of communication, even from “daily” to about “weekly” results in decrease of likelihood of co-authored publication.

The length of the collaborative relationship is universally important predictor of the likelihood of co-authored publication across all models.¹³ Specifically, the likelihood that a collaboration has yielded a co-authored publication in the past 2 years is about 25 % higher in relationships longer than 3 years than in relationships shorter than that. There seem to be “diminishing returns” to length of relationship as the increase in likelihood of co-authored publication in relationships longer than 6 year is not substantially higher than for relationship of 3–6 years (26 vs. 24 percentage points), and the difference is only statistically significant at the 0.10 level).

The survey respondents were also asked to indicate if they consider each individual collaborator to be a “close friend”. Approximately 25 % of all collaborators fell in this category. While there are theoretical reasons to expect positive effect of social

¹² 13 % of all collaborations fall into this category, see Table 2

¹³ Wald test of equivalence of the coefficients on “3–6 years” and “>6 years” suggest that relationships longer than 6 years have larger positive effect on likelihood of co-authored publication than 3–6 year relationships ($P = 0.04$).

relationships on publication productivity, this is not universally supported by the data. While such relationship holds true, it applies in specific contexts: the collaborations of most senior academics (i.e. full professors) and the collaborations of low and medium productivity scholars. These findings reveal interesting dynamics. First, the findings within the context of rank question the traditional assumption that close social relationships facilitate professional success. Indeed, the findings suggest different direction of influence: perhaps it is not that social relationships (e.g., friendship) facilitate collaborative success, but the other way around; maybe after a career long collaborations, some collaborators also become close friends, as the findings within the context of rank would suggest. However, this does not mean that social relationships are inconsequential: they seem to be important for scholars with less robust publication outputs (i.e. low and medium productivity scholars). It should also be noted that this relationship could be somewhat spurious considering that the lower productivity a scholar is, the more likely that he or she has less collaborators, and therefore—the more likely that more of them will be “close friends.”

How the first contact was established seems to have a limited effect on the likelihood of co-authored publication. Specifically, having met for the first time at a conference does not seem to affect the likelihood of co-authored publication, and for the assistant professors and low-productivity individuals it actually hurts the chances of co-authored publication.

Discussion

The findings of this study suggest stable underlying relationships that have evolved over time and are possibly characterized with higher levels of trust are more likely to yield co-authored publications. The implication of this general result is that collaborations are neither discrete nor primarily resource-based or institutionally-influenced. Instead, co-authorship is more likely to be characterized with (1) a pattern of lengthy history amongst co-authors, (2) frequent communication, (3) some level of mutual trust and support extending beyond the direct objects of the collaborations, and (4) shared socialization or educational history.

The findings also suggest that institutional influences may be less important than typically thought. For example, the majority of respondents' close collaborations are with individuals from outside the university, and collaborations with outside individuals tend to be more likely to result in a co-authored publication. Which reinforces the implication that relational history and patterns of trust and communication may be at least as important predictors of co-authorship as other factors more readily tracked with data and thusly more regularly studied (i.e. here institutional factors). However, the boundary-spanning nature of co-authorships this result implies also reinforces that resources like relations matter (insofar that boundary-spanning research is in some cases motivated by resources, see Boardman 2009).

These findings have some preliminary or potential implications for practice as well as for theory.

In terms of policy and managerial implications, the results essentially defy common practice in terms of institutional influences. The findings suggest that institutional attempts to influence scientific collaborations face an uphill battle insofar scientists seem to self-select into collaborations for informal and relational reasons even when confronted with resource-and institution-based incentives. If this is in fact so (pending further investigation, see “[Conclusion](#)” section), then any formal or policy-driven attempts to encourage

collaborations that wouldn't happen otherwise must address in addition to resource needs relational barriers to collaboration (the barrier being the potential lack of these relations). Which suggests that organic research collaborations must not be supplanted by top-down policies but rather facilitated and enhanced by public policy (see Boardman and Bozeman 2006 for further discussion).

First, besides providing incentives for collaboration, institutional attempts to influence collaboration patterns need to provide conditions for sustaining the new collaborations for a long enough period of time so that mutual trust, familiarity, and communication patterns have time to develop. The 5-year cycle of new boundary-spanning organizations established by the NSF may be a guide, however we can't say because we know of no investigation of how long it takes to develop collaborative relationships in science and engineering research.

Second, in trying to facilitate inter-disciplinary research collaborations, institutional attempts to facilitate such collaborations need to perhaps be even more patient, given that bringing researchers from multiple disciplines together by definition means they will not be intimately aware of each other's skills and expertise. Institutional and organizational attempts to influence the patterns of research collaboration should consider specific managerial and human resource mechanisms to create the conditions for effective communication and development of trust and goal congruence across scientists. The budding team science literature is informative in this regard.

In terms of theory, this study has the potential to contribute to the generally neglected issue of validly and reliably operationalizing research collaboration, and to enhance the understanding of structural characteristics of research collaborations. The results undermine the common assumption that collaborations generally materialize in co-authorships, with only about 50 % of the collaborations studied having yielded a co-authored publication, and 24 % producing no bibliometric output—publication, patent, or a conference paper. Related, the fluid content and boundaries of collaborations suggest the any conceptualization of scientific collaborations should expand to include less discrete qualities of the collaboration (such as particular incentives or strategies), and rather turn focus towards understanding research collaboration as an ongoing process. This line of research has been started by the scientific and technical human capital approach (Bozeman et al. 2001).¹⁴

Conclusion

In this paper we've only tried to unravel what co-authorship is as a social phenomenon. Though our findings suggest co-authorship to be as much relational as resource-based, and perhaps less institutionally-influenced than relational, the typical caveats regarding cross-sectional and self-reported data and omitted variable bias apply. Nonetheless the results reveal there's still more to learn about authorship ascription as a relational phenomenon, and this knowledge in turn may help to develop better understanding of research collaboration.

¹⁴ Though the approach suggests already that the scientists' research capacity should incorporate social capital variables, it does not provide guidelines for operationalization. Further, the approach emphasizes social ties as capital or resources, not as relations per se. Accordingly, the approach is in our view inappropriately biased towards the accumulation of institutionally diverse connections, even though existing and homophilic connections are just as important component of co-authorship.

What this means is incorporating what we eventually find about the non-bibliometric aspects of research collaboration—e.g., impetuses, processes, relations, decision calculi for disseminating results via different media—into broader investigations using bibliometric data to analyze the net effects of research policies and programs. Because the current reliance on bibliometric data to operationalize research collaboration that would not have occurred sans a particular public research investment is inaccurate and misleading (Laudel 2002).

This task is challenging both methodologically and empirically. To illustrate, the sociology of science literature has historically shunned bibliometric methods generally because anyone seeking to construct sociological explanations by using bibliometric methods must, from these sociologists' collective view, "cross into a methodological no man's land" (Glaser and Laudel 2001, p. 411). But in our view theory and public policy dictate the challenge must be taken on, however imperfectly.

For us (and perhaps also for you) this means enhancing our survey data and other sources of non-bibliometric information on the relational aspects of research collaboration (e.g., from curriculum vitae and interviews) first with bibliometric data, and next with institutional data. The typical metrics on co-authors' institutional affiliations (which are easily gleaned from bibliometric data) can help to understand the boundary-spanning nature of co-authorship and then hypothesize about the extent to which these boundaries were spanned for resource-based reasons, due to institutions and policy, and/or informal relations. Institutional data will help to move beyond binary measures of the information gleaned from co-authors' institutional affiliations in terms of types and quantities of resources, the formality or lack thereof of institutional norms and expectations, and so on.

The rub then is that, at least to start, research on research collaboration must get smaller not bigger. Most work on research collaboration continues to rely exclusively on bibliometric data, which means going broad rather than deep (Abramo et al. 2009; Butcher and Jeffrey 2005; Glänzel and Schubert 2004; Newman 2004; Wang et al. 2005). Instead deeper investigation must happen not just in the ways we suggest above, but for smaller samples within a single discipline or field. This sort of work in our view is required before broader forays because it's important first to develop an internally valid explanation of the multiple dimensions of co-authorship, much less research collaboration.

Relational as well as resource-based factors (but not institutional ones) explain authorship ascription for this article.

Appendix: Excerpt from the survey questionnaire—Questions and definitions conditioning respondents' answers to collaboration questions

III Your Network

An important focus of our study is on the work relationships that evolve in the science and engineering communities. The following sets of questions ask you about specific types of interactions with people you know. The data collected in these questions is critical for understanding research and other support relationships in the academic environment. *Your completion of this survey is completely confidential and the people you identify will not know that you have named them in a survey.*

The following questions ask you about individuals you have worked and collaborated with at your institution, as well as in other organizations (Collaboration includes proposal

generation, working on a research project, writing/presenting an academic paper/book or book chapter, or developing industrial products or patents). Later, you will be asked additional questions about these individuals.

50. Over the past two academic years, which individuals at your university have been your closest research collaborators? *Example: Chris Smith*
[5 text input fields]

51. Over the past two academic years, who have been your closest research collaborators outside of your institution (including other academic institutions, government or industry?) *Example: Chris Smith*
[5 text input fields]

References

- Abramo, G., D'Angelo, C. A., Di Costa, F., & Solazzi, M. (2009). University-industry collaboration in Italy: A bibliometric examination. *Technovation*, 29(6–7), 498–507. doi:10.1016/j.technovation.2008.11.003.
- Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, 34(3), 259–285. doi:10.1016/j.respol.2005.01.014.
- Adams, J. D., Grant, C., Clemmons, R., & Stephan, P. E. (2003). *Patterns of Research Collaboration in US Universities, 1981–1999*. Paper presented at the AAAS Meetings, Denver, CO.
- Batista, P. D., Campitelli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
- Beaver, D. D. (2001). Reflections on scientific collaboration, (and its study): Past, present, and future. *Scientometrics*, 52(3), 365–377.
- Beaver, D., & Rosen, R. (1978). Studies in scientific collaboration—Part I. The professional origins of scientific co-authorship. *Scientometrics*, 1(1), 65–84.
- Biagioli, M. (2003). Rights or rewards? Changing frameworks of scientific authorship. In M. Biagioli & P. Galison (Eds.), *Scientific authorship: Credit and intellectual property in science*. New York: Routledge.
- Biagioli, M., & Galison, P. (2003). *Scientific authorship: Credit and intellectual property in science*. New York, NY: Routledge.
- Birnholtz, J. P. (2005). *When do researchers collaborate? Toward a model of collaboration propensity in science and engineering research*. Ann Arbor: The University of Michigan.
- Block, F., & Keller, M. R. (2009). Where do innovations come from? Transformations in the US economy, 1970–2006. *Socio-Economic Review*, 7(3), 459–483. doi:10.1093/ser/mwp013.
- Boardman, P. C. (2009). Government centrality to university-industry interactions: University research centers and the industry involvement of academic researchers. *Research Policy*, 38(10), 1505–1516. doi:10.1016/j.respol.2009.09.008.
- Boardman, P. C., & Bozeman, B. (2006). The emergence and impact of 'organic' research collaboration. *Economics of Innovation and New Technology*, 15(1), 51–69.
- Boardman, P. C., & Corley, E. A. (2008). University research centers and the composition of research collaborations. *Research Policy*, 37(5), 900–913. doi:10.1016/j.respol.2008.01.012.
- Boardman, P. C., & Ponomariov, B. L. (2009). University researchers working with private companies. *Technovation*, 29(2), 142–153. doi:10.1016/j.technovation.2008.03.008.
- Bordons, M., & Gomez, I. (2000). *Collaboration networks in science web of knowledge—A Festschrift in honor of Eugene Garfield* (pp. 197–213). Medford: Information Today Inc.
- Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy*, 33(4), 599–616.
- Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: An alternative model for research evaluation. *International Journal of Technology Management*, 22(7–8), 716–740.
- Bozeman, B., Fay, D., & Slade, C. P. (2012). Research collaboration in universities and academic entrepreneurship: The-state-of-the-art. *The Journal of Technology Transfer*, 38(1), 1–67. doi:10.1007/s10961-012-9281-8.

- Bozeman, B., & Youtie, J. (2015). Trouble in paradise: Problems in academic research co-authoring. *Science and Engineering Ethics*. doi:10.1007/s11948-015-9722-5.
- Butcher, J., & Jeffrey, P. (2005). The use of bibliometric indicators to explore industry–academia collaboration trends over time in the field of membrane use for water treatment. *Technovation*, 25(11), 1273–1280. doi:10.1016/j.technovation.2004.06.003.
- Corley, E., Boardman, P. C., & Bozeman, B. (2006). Design and the management of multi-institutional research collaborations: Theoretical Implications from two case studies. *Research Policy*, 35(7), 975–993.
- Drenth, J. H. (1998). Multiple authorship: The contribution of senior authors. *JAMA*, 280(3), 219–221. doi:10.1001/jama.280.3.219.
- Flanagin, A., Fontanarosa, P. B., & DeAngelis, C. D. (2002). Authorship for research groups. *JAMA*, 288(24), 3166–3168. doi:10.1001/jama.288.24.3166.
- Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In H. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). Berlin: Springer.
- Glaser, J., & Laudel, G. (2001). Integrating scientometric indicators into sociological studies: methodical and methodological problems. *Scientometrics*, 52(3), 411–434.
- Gordon, M. (1980). A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics*, 2, 193–201.
- Guston, D. H. (2000). *Between politics and science: Assuring the integrity and productivity of research*. Cambridge: Cambridge University Press.
- Hagedoorn, J., Link, A. N., & Vonortas, N. S. (2000). Research partnerships. *Research Policy*, 29(4–5), 567–586. doi:10.1016/S0048-7333(99)00090-6.
- Hagstrom, W. O. (1965). *The scientific community*. New York: Basic Books.
- Harsanyi, M. A. (1993). Multiple authors, multiple problems bibliometrics and the study of scholarly collaboration—A literature-review. *Library & Information Science Research*, 15(4), 325–354.
- Jeong, S., Choi, J., & Kim, J. (2011). The determinants of research collaboration modes: Exploring the effects of research and researcher characteristics on co-authorship. *Scientometrics*, 89(3), 967–983. doi:10.1007/s11192-011-0474-y.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1–18.
- Kingsley, Gordon, Bozeman, Barry, & Coker, Karen. (1996). Technology transfer and absorption: An ‘R & D value-mapping’ approach to evaluation. *Research Policy*, 25(6), 967–995.
- Kreiner, K., & Schultz, M. (1993). Informal collaboration in research-and-development—The formation of networks across organizations. *Organization Studies*, 14(2), 189–209.
- Laudel, G. (2001). Collaboration, creativity and rewards: why and how scientists collaborate. *International Journal of Technology Management*, 22(7–8), 762–781.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3–15.
- Lundberg, G. D., & Glass, R. M. (1996). What does authorship mean in a peer-reviewed medical journal? *JAMA*, 276(1), 75. doi:10.1001/jama.1996.03540010077036.
- Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., & Brommels, M. (2006). Collaboration uncovered: Exploring the adequacy of measuring university–industry collaboration through co-authorship and funding. *Scientometrics*, 69(3), 575–589. doi:10.1007/s11192-006-0170-5.
- Melin, G. (2000). Pragmatism and self-organization—Research collaboration on the individual level. *Research Policy*, 29(1), 31–40.
- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363–377.
- Mullins, N. C. (1973). *Science: Some sociological perspectives*. Indianapolis: Bobbs-Merrill.
- Narin, F., & Whitlow, E. S. (1991). *Measurement of scientific cooperation and coauthorship in CEC-related areas of science*: Commission of the European Communities Directorate-General Telecommunications. Luxembourg: Information Industries and Innovation.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(Suppl 1), 5200–5205.
- Ponomariov, B., & Boardman, P. C. (2008). The effect of informal industry contacts on the time university scientists allocate to collaborative research with industry. *Journal of Technology Transfer*, 33(3), 301–313. doi:10.1007/s10961-007-9029-z.
- Ponomariov, B. L., & Boardman, P. C. (2010). Influencing scientists’ collaboration and productivity patterns through new institutions: University research centers and scientific and technical human capital. *Research Policy*, 39(5), 613–624. doi:10.1016/j.respol.2010.02.013.

- Porter, A. L., Roessner, J. D., Cohen, A. S., & Perreault, M. (2006). Interdisciplinary research: meaning, metrics and nurture. *Research Evaluation*, 15(3), 187–195. doi:[10.3152/147154406781775841](https://doi.org/10.3152/147154406781775841).
- Price, D. J. D., & Beaver, D. D. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011.
- Qiu, L. (1992). A study of interdisciplinary research collaboration. *Research Evaluation*, 2(3), 169–175. doi:[10.1093/rev/2.3.169](https://doi.org/10.1093/rev/2.3.169).
- Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable. *JAMA*, 278(7), 579–585. doi:[10.1001/jama.1997.03550070071041](https://doi.org/10.1001/jama.1997.03550070071041).
- Riesenberg, D., & Lundberg, G. D. (1990). The order of authorship: Who's on first? *JAMA*, 264(14), 1857. doi:[10.1001/jama.1990.03450140079039](https://doi.org/10.1001/jama.1990.03450140079039).
- Shapiro, D. W., Wenger, N. S., & Shapiro, M. F. (1994). The contributions of authors to multiauthored biomedical research papers. *JAMA*, 271(6), 438–442. doi:[10.1001/jama.1994.03510300044036](https://doi.org/10.1001/jama.1994.03510300044036).
- Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, 41(1), 643–681. doi:[10.1002/aris.2007.1440410121](https://doi.org/10.1002/aris.2007.1440410121).
- Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal Of Information Science*, 6(1), 33–38. doi:[10.1177/016555158300600105](https://doi.org/10.1177/016555158300600105).
- Traoré, N., & Landry, R. (1997). On the determinants of scientists' collaboration. *Science Communication*, 19(2), 124–140. doi:[10.1177/1075547097019002002](https://doi.org/10.1177/1075547097019002002).
- Van Raan, A. (1998). The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations. *Scientometrics*, 42(3), 423–428.
- Wang, Y., Wu, Y., Pan, Y., Ma, Z., & Rousseau, R. (2005). Scientific collaboration in China as reflected in co-authorship. *Scientometrics*, 62(2), 183–198.
- Wray, K. B. (2002). The epistemic significance of collaborative research. *Philosophy of Science*, 69(1), 150–168. doi:[10.1086/338946](https://doi.org/10.1086/338946).
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
- Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics*, 8(2), 295–309. doi:[10.1016/j.joi.2014.01.008](https://doi.org/10.1016/j.joi.2014.01.008).
- Yoshikane, F., & Kageura, K. (2004). Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, 60(3), 433–444.