CREMA
Center for Research in Economics, Management and the Arts

The First Cut is the Deepest: Repeated Interactions of Coauthorship and Academic Productivity in Nobel Laureate Teams

Working Paper No. 2015-04

CREMA Südstrasse 11 CH - 8008 Zürich www.crema-research.ch

The First Cut is the Deepest: Repeated Interactions of Coauthorship and Academic Productivity in Nobel Laureate Teams

Ho Fai Chan[a], Ali Sina Önder[b], & Benno Torgler[a,c,*]

[a] *Queensland University of Technology, School of Economics and Finance, QuBE, 2 George St, Brisbane QLD 4000, Australia.*
[b] *University of Bayreuth, Department of Economics, Universitätsstrasse 30, D-95447 Bayreuth, Germany.*
[c] *CREMA, Südstrasse 11, CH-8008 Zürich, Switzerland.*

Abstract: Despite much in-depth investigation of factors influencing this evolution in various scientific fields, our knowledge about how efficiency or creativity is linked to the longevity of collaborative relationships remains very limited. We explore what Nobel laureates' coauthorship patterns reveal about the nature of scientific collaborations looking at the intensity and success of scientific collaborations across fields and across laureates' collaborative lifecycles in physics, chemistry, and physiology/medicine. We find that more collaboration with the same researcher is actually no better for advancing creativity: publications produced early in a sequence of repeated collaborations with a given coauthor tend to be published better and cited more than papers that come later in the collaboration with the same coauthor. Thus, our results indicate that scientific collaboration involves conceptual complementarities that may erode over a sequence of repeated interactions.

JEL classification: D20; O30

Keywords: Innovation; Scientific Collaboration; Team Formation; Nobel Laureates

---

[*] Corresponding author. Email: benno.torgler@qut.edu.au (B. Torgler).

1. Introduction

Dramatic changes in science over the past decades have increased task complexity, reshaping how scientists cooperate and turning science into a team effort (Katz & Martin, 1997; Adams et al. 2005). In particular, the one-author-per-paper trend that dominated science from the 1600s until around the 1920s decreased in the 1950s, was barely visible by the 1980s (Greene, 2007), and has become a rarity in scientific journals today. For example, of the 700 reports published in *Nature* in the first 10 months of 2008, only six were single author papers (Whitfield, 2008). Our understanding of such collaboration is informed by visualisation of collaborative patterns (Newman, 2004) and an evolving understanding of the principles of team formation (Guimera et al., 2005; Milojević, 2014), which provides useful insights into optimal team size. The emerging use by scientists of collaborative indexes to more effectively measure researchers' scientific impact (Stallings et al., 2013) also suggests that in the past few decades, single authors have performed worse than teams (Wuchty et al., 2007). Nevertheless, knowledge of how teams perform over time remains limited.

To help fill this void, we explore the productivity patterns of repeated scientific collaborations by Nobel laureates and their collaborators, thus ensuring a homogenous focus group of productive scientific "stars" with intellectual human capital of extraordinary scientific value. In particular, laureates are homogenous in their capacity to produce successful, innovative ideas and attract fairly able coauthors (Zuckerman, 1996), which allows us to focus on team efficiency while holding team talent constant.

2. Data and Descriptive Analysis

Our dataset consists of 31,832 publications registered in Scopus (up to 2008) of 192 Nobel laureates who received the Nobel Prize in chemistry (56), physics (69), or physiology/medicine (67) between 1970 and 2000. The dataset includes 42,222 Nobel laureate coauthor pairs, for whose publications citation records are traceable up to 2014. The patterns of laureates' accumulation of coauthors are similar in different fields. Although most Nobel laureates cooperate with fewer than 160 different coauthors over their academic lifecycle, a few cooperate with over 1,000 different coauthors. The long tails of the histograms (Figure A1) somewhat reflect the fact that "hyper-authorships" tend to be the product of highly complex subfields such as biomedicine or high-energy physics (Cronin, 2001).

Our first analysis explores the arrival of new coauthors and the intensity of coauthorship over Nobel laureates' academic lifecycle. Figure 1 shows the number of new coauthors that appear in laureates' publications at a given age. The patterns for the arrival of new coauthors are comparable in chemistry, physics, and physiology/medicine before laureates reach age 60. Age 60 marks the peak for arrival of new coauthors in chemistry and physics, although there seems to be no clear peak in physiology/medicine.

The intensity of coauthorship captures the number of total collaborations between a laureate and a given coauthor (Figure 2). "Laureates' Age" corresponds to the time of first collaboration, with the vertical axis depicting the average number of collaborations between Nobel laureates and arriving coauthors (i.e., when collaboration begins) at a given age for the laureate. In chemistry and medicine, early collaborations tend to be more intense (albeit with a large variance). In physics, however, laureates' intensities of coauthorship tend to show a positive trend at younger ages but no clear peak is observed.

We choose arrival of new coauthors and intensity of collaboration to capture the dynamics of Nobel laureates' collaborations over their academic lifecycle because they reflect the social and academic norms in the respective fields. We assess the quality of such collaborations based on the number of citations received. For every laureate-coauthor pair that has published collaboratively in at least 4 distinct years, we calculate the average number of citations received by publications during first 2 years and last 2 years of collaboration. Figure 3 then plots the differences between the two publication sets (first 2 years minus last 2 years) on the horizontal axis against the average number of citations received by publications in first 2 years on the vertical axis (panel a). Panel b contains data restricted to laureate-coauthor pair that has published collaboratively in at least 7 distinct years. The green diagonal line indicates early-late citations differences are equal to early citations. Results reveal that collaboration success is minimally dependent on pure luck: laureate and coauthor pairs that receive a high number of citations for their later publications are also those who receive a high number for their early publications. Conversely, most collaborations that yield no highly cited publications early on tend to yield even fewer successful publications down the road.

The decay in citation success appears to be strongest in chemistry, in which many observations fall very close to the point at which the difference between early and late citations is equal to early citations (hence zero late citations), designated by the green diagonal line. Collaborations in physics and physiology/medicine that repeat over many years show more success than those in chemistry, in which most observations are placed very close to the diagonal line and to the right of the vertical line. Observations in physics and physiology/medicine, on the other hand, fall further away from the diagonal line and to the left of the vertical line, which hints at not "too unsuccessful" or even more successful late collaborations between laureate-coauthor pairs. The differences in citation success in earlier versus later publications over the lifecycle of a given collaboration is greater in chemistry, perhaps because most chemistry research is done in a way to generate very specific data that are best published within a few high impact publications. Research in physics and physiology/medicine, on the other hand, generate rather more multidimensional data that sustain a large number of good ideas leading to several high impact publications, especially in highly complex research areas where experiments require a very costly setup.

3. Two-Stage Estimation and Discussion of Results

In the first stage, to isolate the correlation between citations received for an article and the intensity of cooperation between that article's coauthors, we define journal quality as the journal's 2012 impact factor from the ISI Web of Knowledge 2012 Journal Citation Reports and regress this variable on paper characteristics in the first stage estimation to obtain prediction errors ($\hat{\mu}_{ij,h}$). In the second stage estimation, we regress citation count on the same explanatory variables as in the first stage but also on the predicted errors derived therein. The journal impact factor in the second step is thus the error obtained in the first, corresponding to the portion of journal impact factor not explained by the paper and collaboration characteristics. In this way, we separate the effects of journal quality on citation success from other explanatory variables.

The bases for these estimations are the following two specifications:

$$\begin{aligned}&\text{Step 1: } (Journal\_Impact)_{ij,h}\\&= f(total\ collaboration, collaboration\ year, \#authors, laureate\ characteristics)\\&+ \mu_{ij,h}\end{aligned}$$

$$\begin{aligned}&\text{Step 2: } (Citations)_{ij,h}\\&= f\big(total\ collaboration, collaboration\ year, \#authors, laureate\ characteristics, \hat{\mu}_{ij,h}\big)\\&+ \varepsilon_{ij,h}\end{aligned}$$

We regress the journal impact factor for paper *h* of the laureate-coauthor pair *ij* on the total number of laureate (*i*) and coauthor (*j*) collaborations in our dataset (*total collaboration*, different dummy variables), the year of appearance of that particular paper *h* in the life cycle of *ij* collaboration (in the first year, second year, or *n*th year of the collaboration), the total number of authors in publication *h*, and the Nobel laureate's characteristics (field, age during publication, and individual fixed effects). To avoid collinearity between the total number of collaborations and the appearance number of a particular collaboration, we use indicator variables for various levels of total collaboration: 6 to 20, 21 to 40, 41 to 70, 71 to 110, and more than 110 (with between 1 and 5 as the reference group).

We focus on the marginal effects of repeated collaborations between laureate-coauthor pairs on citation success of their publications. In doing so, we must recognise that citations may be affected by the quality of the journal in which the article is published (e.g., due to increased visibility), meaning that when citations are regressed on paper characteristics that include publishing journal quality (measured by impact factor), such quality will be highly correlated with other explanatory variables. This correlation could produce misleading outcomes because journal quality and citation of the article, rather than being independent, are determined by the same exogenous factors, including collaboration intensity. The citation success results show that the first four collaboration bins are all highly significant but negative (relative to the reference group of 5 or fewer collaborations), with only the fifth bin, the most extreme number of collaborations, being positive and significant (Table 1). Hence, all else being equal, and except for the extreme case of over 110 collaborations, the first cooperation sets tend to be more successful, leading to more citations per paper (between 19 and 54). Among laureates who won the prize while under 50, collaborations repeating more than 20 times have a positive and significant coefficient. For the laureates who won the prize after 50, the most successful papers are the early publications with the most intensive collaboration (over 110 repeated interactions). Most laureate-coauthor pairs collaborate over several years. The year (e.g., first, second, third …) of the laureate-coauthor collaboration in which a particular publication occurs is captured by the variable *Collaboration Year* in table 1. Square of the collaboration year is included to capture the non-linear productivity pattern over the life cycle of collaborations. Long lasting collaborations are those that produce as good (or even better) cited publications during later years of collaboration as in the early years of it, and this is revealed by the non-linear marginal effect of the *collaboration year.* Non-linearity of citation success over the life cycle of a given collaboration captures an interesting relationship: although creativity and impact decays over the life cycle of many collaborations (most repeat over less than 4 distinct years), there are also some very long lasting collaborations that don't experience such a strong decay in productivity.

Comparing our results for different fields, we find that although the total number of citations received by a paper in chemistry and medicine is strongly positively correlated with the total number of collaborations between the laureate and that particular coauthor, earlier papers in the collaboration sequence are expected to receive higher citations. In physics, on the other hand, total number of citations is strongly negatively correlated with total collaborations, except for collaborations that repeat more than 110 times, thus most citations are expected for papers from collaborations that repeat either less than 5 times or more than 110 times.

What, then, are the most likely reasons for our results, which suggest that, based on citation count, the first set of ideas generated with a coauthor is the most innovative, and that a collaboration may run out of creative ideas over time ("collaborative idea scarcity"). One explanation may be that the creativity of the original combination that generates new insights and breakthroughs may emerge early in the researchers' collaboration. Likewise, efficient problem solving may emerge initially but become less relevant after success is achieved. From then on, the pool of creative ideas seems to decrease. These assumptions are somewhat supported by the evidence that success may be augmented by pairing high conventionality with novelty using atypical combinations (Uzzi et al., 2013) that themselves may be encouraged by novel interactions. It is also possible that highly innovative researchers such as Nobel laureates may be more critical of new collaborations and may only agree to those that seem to offer meritorious rigor. On the other hand, collaborations may be chosen for reasons other than their effect on output or may, over time, transform into friendships, which decreases the pressure to collaborate productively (Hollis, 2001). That is, cooperation can lead to an intellectual companionship that overcomes isolation, creating a personal relationship between the coauthors (Katz and Martin, 1997). Thus, whereas a new collaboration can enhance diversity of perspective, more intensive collaboration may reduce diversity not only in perspective but also in expertise and experience. It is also important to consider the environment in which research is being produced. It is common in natural sciences that large research teams are gathered together around specific topics so that scientific investigations or experiments incur large monetary costs. In such complex scientific research, the collaborative lifespan, as well as the richness of publishable material (e.g., continuing stream of data and results) may mostly be prescribed by the initial setting.

4. Conclusion

One definite strength of new collaborations is that they are often characterised by a willingness to consider new ideas and/or adapt to novel approaches. In any collaboration—but particularly in science—trust is crucial to the sharing of ideas, models, data, or material of substantial scientific merit; hence, the scientific colleagues of Nobel laureates may be more willing to trust Nobellists in that regard. Hence, to benefit from the increasingly collaborative nature of scientific inquiry, researchers need a better understanding of what determines team success. The results reported here suggest that the advantages and costs of ongoing collaboration should be carefully weighed because, from a creativity viewpoint, collaborations have an expiration date even for Nobel laureates.

References

Adams JD, Black GC, Clemmons JR and Stephan PE (2005) Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999. *Research Policy* 34: 259–285.

Cronin B (2001) Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology* 52(7): 558–569.

Greene M (2007) The demise of the lone author. *Nature* 450: 1165.

Guimera R, Uzzi B, Spiro J and Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308: 697–702.

Hollis A (2001) Co-authorship and the output of academic economists. *Labour Economics* 8(4): 503–530.

Katz JS and Martin BR (1997) What is research collaboration? *Research Policy* 26: 1–18.

Milojević S (2014) Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 111: 3984–3989.

Newman MEJ (2004) Who is the best connected scientist? A study of scientific coauthorship networks. In: Ben-Naim E, Frauenfelder H, Toroczkai Z (eds) *Complex Networks*. Berlin: Springer, 337–370.

Stallings J, Vance E, Yang J, Vannier MW, Liang J, Pang L, et al. (2013) Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences of the United States of America* 110: 9680–9685.

Uzzi B, Mukherjee S, Stringer M and Jones B (2013) Atypical combinations and scientific impact. *Science* 342: 468–472.

Whitfield J (2008) Group theory. *Nature* 455: 720-723.

Wuchty S, Jones BF and Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316: 1036–1039.

Zuckerman H (1996) *Scientific Elite: Nobel Laureates in the United States*. New Brunswick: Transaction Publishers.

Tables and Figures

Table 1. Regression results for the 2SLS

	Overall	**Age of the Laureate**		**Chemistry**	**Physics**	**Phy./Med**
		<50	**>50**			
Collaborations 6-20	-18.96***	-18.54**	2.76	27.26***	-16.05***	29.21***
	(2.452)	(7.478)	(2.055)	(4.786)	(2.522)	(3.800)
Collaborations 21-40	-24.50***	49.75***	9.85***	43.72***	-16.64***	87.17***
	(2.775)	(9.189)	(2.496)	(8.343)	(2.753)	(5.506)
Collaborations 41-70	-53.89***	47.48***	-3.22	35.00***	-21.18***	86.89***
	(3.364)	(9.001)	(2.627)	(8.665)	(2.876)	(7.019)
Collaborations 71-110	-38.14***	64.35***	3.92	43.65***	-18.73***	100.85***
	(3.483)	(11.054)	(2.856)	(16.429)	(2.858)	(7.961)
Collaborations >110	6.90*	54.71***	14.78***	26.19***	3.57	110.23***
	(3.529)	(13.581)	(3.156)	(7.824)	(2.963)	(8.005)
Collaboration Year	-2.96***	-11.93***	-3.48***	-7.14***	-3.16***	-12.11***
	(0.485)	(1.661)	(0.466)	(1.289)	(0.539)	(1.041)
Collaboration Year^2	0.17***	0.38***	0.14***	0.27***	0.19***	0.39***
	(0.019)	(0.061)	(0.019)	(0.045)	(0.036)	(0.038)
Journal Quality	8.81***	8.47***	5.81***	11.52***	5.80***	3.84***
	(0.227)	(0.424)	(0.224)	(0.682)	(0.472)	(0.185)
Number of Authors	2.31***	6.92***	0.32***	7.66***	0.38***	22.07***
	(0.076)	(0.335)	(0.018)	(0.451)	(0.011)	(0.511)
Laureate Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	144,531	26,412	102,579	30,530	71,561	42,440
Adjusted R-squared	0.36	0.46	0.17	0.18	0.22	0.67

Notes: The table reports second stage coefficients where the dependent variable is an article's citation count. We run our analysis first for all Nobel laureates and then separately for laureates who won the prize while under or over the age of 50. We then conduct regressions for each field separately. Robust standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.

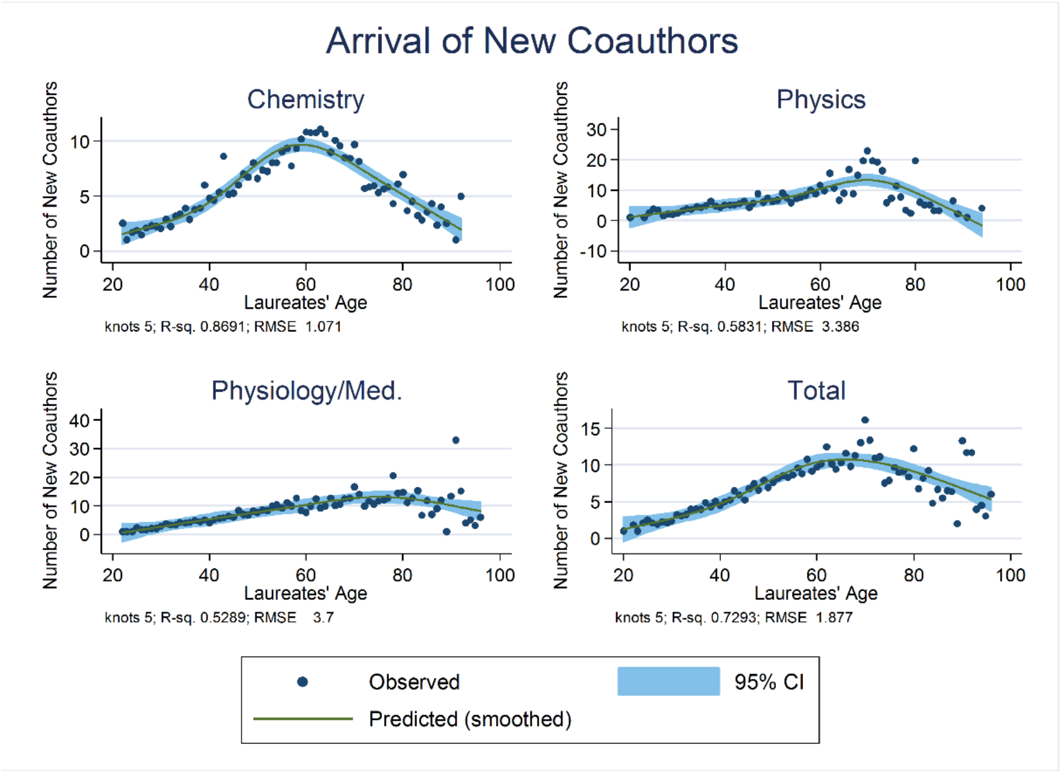

Figure 1. Arrival of new coauthors by field

Note: Smoothed values are computed using restricted cubic spline.

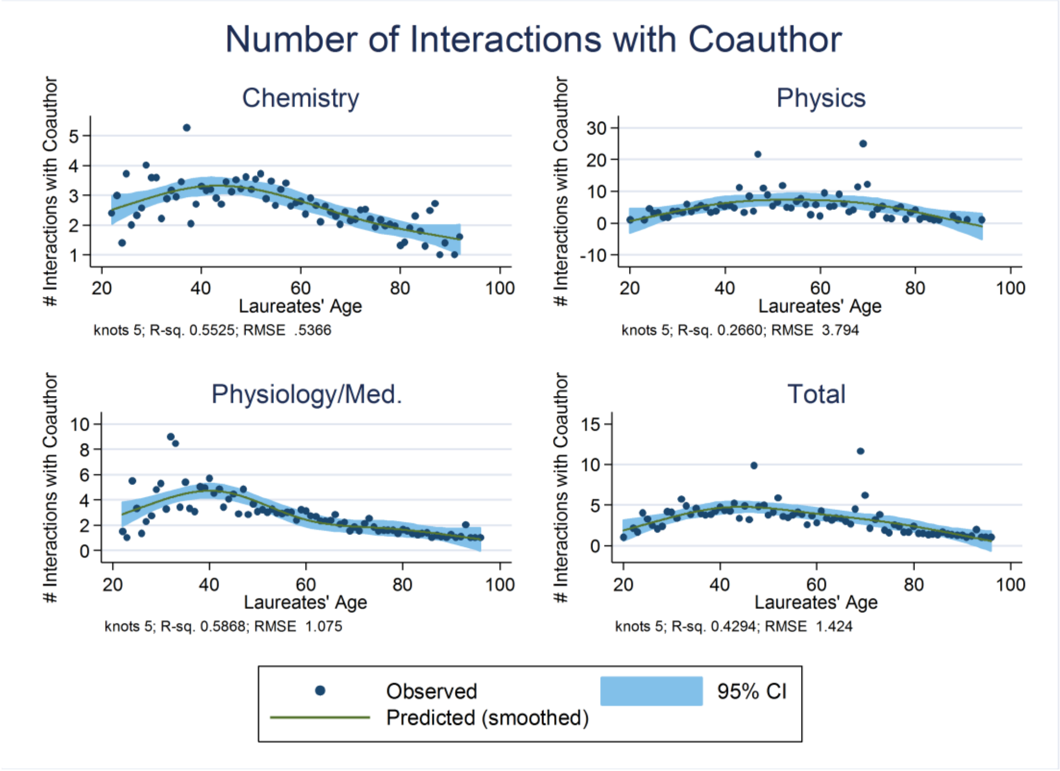

Figure 2. Intensity of cooperation by field

Note: Smoothed values are computed using restricted cubic spline.

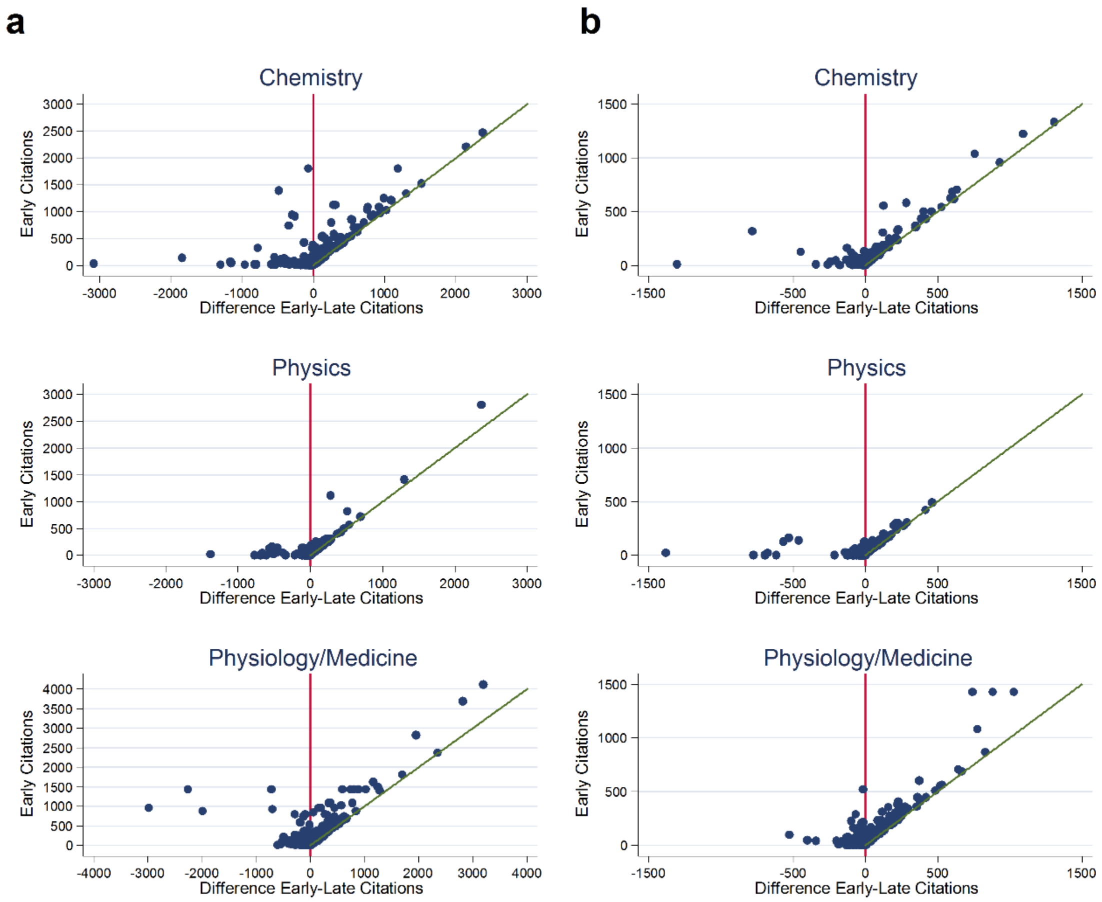

Figure 3. Citations received by early and late collaborations of laureate-coauthor pairs

Appendix

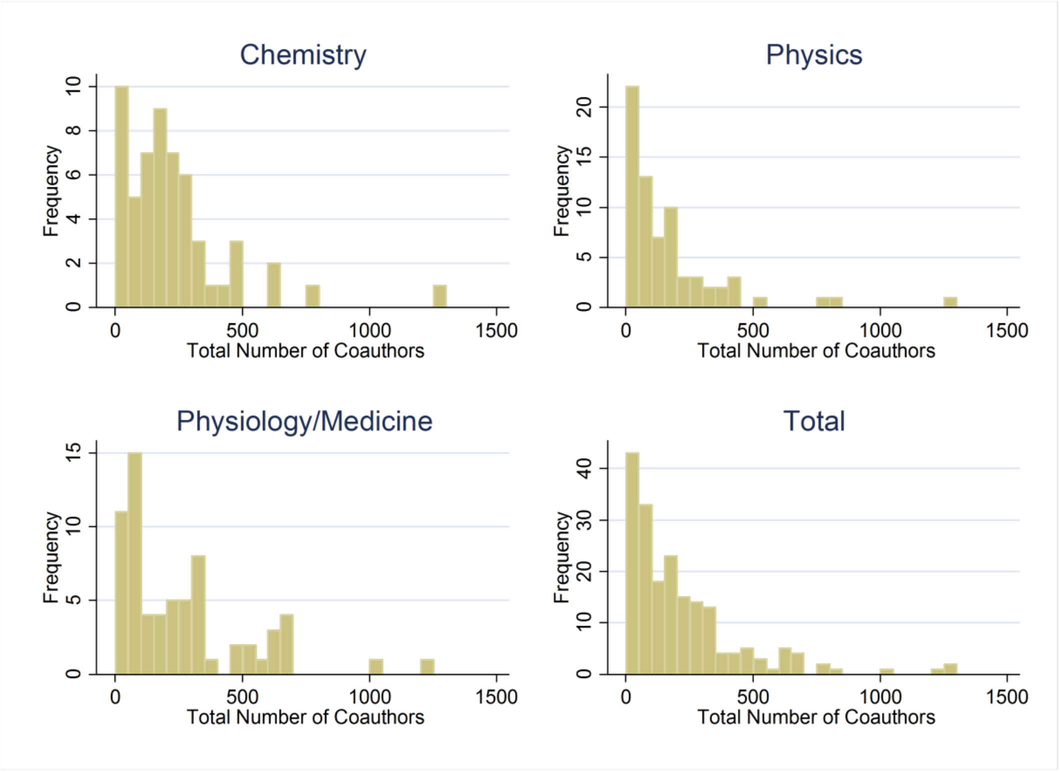

Figure A1. Distribution of the total number of Nobel laureate coauthors.
Note: Bin width = 50.

Table A1. Descriptive Statistics dependent and independent variables employed in 2SLS regression analysis.

	Mean	Mode	Std. Dev.	Min	Max
Citations Received	85.39	18	409.06	1	8947
Journal Quality (Impact Factor)	7.30	4.69	7.63	0.17	51.66
Collaborations 6-20	0.27	0	0.44	0	1
Collaborations 21-40	0.16	0	0.36	0	1
Collaborations 41-70	0.12	0	0.33	0	1
Collaborations 71-110	0.06	0	0.24	0	1
Collaborations >110	0.04	0	0.19	0	1
Collaboration Year	3.61	2	3.80	1	36
Number of Authors	52.37	9	65.64	1	181

Table 2

First Stage Regression Results for 2SLS

	Overall	**Age of the Laureate**		**Chemistry**	**Physics**	**Phy./Med**
		<50	**>50**			
Collaborations 6-20	0.20***	0.82***	0.23***	0.32***	0.37***	0.59***
	(0.058)	(0.142)	(0.069)	(0.101)	(0.057)	(0.133)
Collaborations 21-40	0.38***	1.38***	0.42***	0.97***	0.75***	0.16
	(0.068)	(0.197)	(0.081)	(0.191)	(0.056)	(0.204)
Collaborations 41-70	0.46***	1.20***	0.57***	0.93***	0.74***	1.13***
	(0.079)	(0.191)	(0.099)	(0.223)	(0.058)	(0.328)
Collaborations 71-110	0.68***	1.68***	0.83***	0.85**	0.78***	1.80***
	(0.099)	(0.222)	(0.121)	(0.380)	(0.062)	(0.378)
Collaborations >110	0.76***	1.32***	0.89***	1.36***	0.79***	1.30***
	(0.138)	(0.395)	(0.156)	(0.252)	(0.070)	(0.353)
Collaboration Year	-0.16***	-0.22***	-0.20***	-0.20***	-0.13***	-0.26***
	(0.013)	(0.033)	(0.017)	(0.031)	(0.011)	(0.035)
Collaboration Year^2	0.00***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of Authors	0.02***	0.02***	0.01***	0.18***	0.01***	0.10***
	(0.000)	(0.002)	(0.001)	(0.015)	(0.000)	(0.003)
Laureate Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	144,531	26,412	102,579	30,530	71,561	42,440
Adjusted R-squared	0.26	0.40	0.15	0.13	0.30	0.20

Notes: First stage coefficients are being reported, where dependent variable is the impact factor of the journal where the article is published. Robust standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.