Paper to be presented at

DRUID15, Rome, June 15-17, 2015

(Coorganized with LUISS)

The Anatomy of Teams: Division of Labor in Collaborative Knowledge Production

Carolin Haeussler
University of Passau
Chair of Organization, Technology Management and Entreprene
carolin.haeussler@uni-passau.de

Henry Sauermann
Georgia Institute of Technology
Scheller College of Business
henry.sauermann@scheller.gatech.edu

Abstract

Teams have become increasingly important in knowledge production, yet how teams divide tasks among their members remains ill-understood. Complementing recent work that views innovation as the recombination of prior knowledge, we conceptualize knowledge production as a process involving a number of functional activities and develop a theoretical framework to study the division of labor in scientific teams. We use this framework to examine division of labor empirically using novel data on the activities of all authors who contributed to over 13,000 scientific articles. We find that division of labor is stronger with respect to some functional activities than others, likely reflecting differences in the benefits from specialization and the interdependencies between activities. Division of labor increases with team size, although at a diminishing rate. Moreover, while the share of members performing empirical activities is largely stable across the team size distribution, the share of members engaged in conceptual activities declines sharply, suggesting that conceptual activities may benefit less from parallel processing by multiple team members. We consider implications for research on the organization of knowledge production as well as practical implications for scientists and policy makers.

Jelcodes:O31,-

The Anatomy of Teams:
Division of Labor in Collaborative Knowledge Production

1 Introduction

A growing body of research examines the organization of knowledge production and the role of collaboration in teams. The dominant approach conceptualizes innovation as the recombination of prior pieces of knowledge (Simonton, 2003). That line of work shows how the collaboration of specialized researchers with knowledge in different domains may allow teams to push deeper towards the knowledge frontier and to create more novel combinations of knowledge (Jones, 2009; Singh & Fleming, 2010; Bercovitz & Feldman, 2011; Teodoridis, 2013). However, scientific research involves more than the recombination of different pieces of knowledge. Qualitative descriptions suggest that research is a time-consuming process that involves a number of different functional steps, ranging from the generation of research hypotheses and the design of experiments to the execution of empirical work and the "writing up" of research results in a scientific paper (Latour & Woolgar, 1979; Knorr-Cetina, 1999; Owen-Smith, 2001). A conceptualization of science as a process is also implicit in recent portrayals of the "lab as a firm", which suggest that scientific competition leads labs to pursue efficient knowledge production by relying on the division of labor among specialized team members (Hackett, 1990; Freeman et al., 2001; Stephan, 2012). Despite rich qualitative discussions, however, we have a limited understanding of the process of knowledge production in teams and, in particular, of the division of labor among collaborators. Empirically, there is little large-scale evidence regarding whether and how different activities are divided among team members.[1] More fundamentally, it is not clear whether and how existing organizational theories originally developed to understand the production of physical goods (Smith, 1776) can usefully be applied in studying the division of labor in the production of knowledge.

This paper makes two main contributions. First, we provide conceptual insights into different aspects of division of labor in teams by developing a framework that allows us to consider three complementary perspectives: Focusing on individuals and looking across activities, the first perspective informs us about the degree to which members specialize in a small number of activities vs. perform a broader range of activities. Focusing on activities and looking across team members, the second perspective informs us about the degree to which an activity is distributed across a large share of team members vs. concentrated among a small share. Looking across both individuals and activities, the third perspective informs us about the relationships among activities, e.g., which activities tend to be performed jointly by the same individuals. We then draw on organizational theories to provide a more detailed

[1] A recent exception is (Shibayama et al., 2013) who gain insights into task allocation by surveying corresponding authors of a sample of Japanese publications.

discussion of potential drivers of different aspects of division of labor, considering a number of mechanisms including benefits from specialization in an activity, interdependencies among activities and resulting coordination costs, the decomposability of a given activity, and the overall size of the project. Overall, while prior work has focused on certain aspects of division of labor or a limited number of mechanisms, our framework provides an integrated view that may provide a useful basis for future work on the division of labor in teams.

Second, we contribute to the theoretical and empirical understanding of the organization of knowledge production in teams. Drawing on our framework and prior research in the sociology and economics of science, we first develop conjectures regarding the nature of key aspects of division of labor in the particular context of scientific knowledge production. We then exploit a new data source to provide unique empirical insights into division of labor in scientific teams. In particular, an increasing number of journals require that articles disclose which particular authors have made which types of contributions. We analyze data from over 13,000 such articles - mostly in the biological sciences and medicine - to gain insights into the types of contributions made by each co-author. Our analysis reveals a number of findings. First, while division of labor can be summarized using team-level measures, important additional insights can be gained using measures at the level of individual scientists or activities. For example, while some activities such as experimental work or the supply of materials are often performed by specialized individuals, others – such as data analysis and writing the paper – are more commonly performed jointly with other activities, consistent with differences in the benefits from specialization or in interdependencies between tasks. Consistent with prior theory, we also find that division of labor increases with team size. Again, however, there is significant heterogeneity across activities. While larger teams involve a correspondingly larger number of individuals in empirical activities, the number of members engaged in conceptual activities or in writing grows much more slowly with team size, leading to a sharp drop in the share of team members engaged in such activities in larger teams. This result suggests that some activities can more readily benefit from parallel work of team members than others, perhaps reflecting challenges in coordinating and integrating individual contributions but also systematic differences in the volume of work to be done. To complement our analysis of these basic patterns of division of labor, we conduct a number of exploratory analyses. Among others, we examine how division of labor relates to characteristics of scientific fields (e.g., novelty), attributes of specific projects (e.g., interdisciplinarity), and to the institutional context (e.g., industry vs. academia).

Our insights contribute to a growing literature on the production of knowledge and the economics of science (Dasgupta & David, 1994; Wuchty et al., 2007; Singh & Fleming, 2010; Jones & Weinberg, 2011; Stephan, 2012). Complementing the common focus on collaborators as contributors of different pieces of knowledge, we study their role in performing different types of activities. While activities clearly draw on knowledge as well, the division of labor with respect to activities can differ significantly from the partitioning of knowledge (Postrel, 2002; Takeishi, 2002)[2] and focusing on activities allows us to provide novel insights into the actual research process. Our conceptual discussion and empirical results also show how existing theories of organization can be leveraged to guide studies of the organization of knowledge production.

By examining the organization of collaborative work in knowledge production, we also contribute to the broader organizational literature. Starting from the early work by Smith and others, much of the literature has focused implicitly or explicitly on the production of physical goods in the context of industrial organizations (Smith, 1776; Hamilton et al., 2003). We consider how some of the fundamental concepts translate into the increasingly important context of team-based knowledge production. At the same time, our results also suggest important areas for future research and conceptual broadening. For example, while conventional models of division of labor focus on the trade-off between benefits from specialization and coordination costs, knowledge production may also benefit from the diversity associated with multiple members being engaged in the same activity (Van Knippenberg et al., 2004), potentially offsetting returns from specialization. Similarly, it may be useful to complement the conventional focus on efficiency in production with greater attention to other organizational goals such as team members' learning and professional development, an issue that becomes particularly salient when we think about division of labor in the context of academic science (Hackett, 1990).

Finally, our conceptual framework and empirical results may also be useful for scientists as well as for policy makers. For example, insights regarding the degree to which scientists specialize in certain activities may inform current debates concerning the content and process of science education (Freeman et al., 2001; Agarwal & Sonka, 2010; Fuhrmann et al., 2011). Similarly, thinking about factors that enable or constrain division of labor may help scientists as they consider the benefits or costs of forming larger teams (Gans & Murray, 2013) or as they explore new forms of organizing science such as the crowdsourcing of certain parts of the research process (Nielsen, 2011; Franzoni & Sauermann, 2014).

2 General Framework

2.1 Conceptualizing division of labor in teams

The study of division of labor has a long tradition in economics and organizational research. However, the term has been defined in different ways and applied to different levels of analysis including societies, industries, and individual organizations (Smith, 1776; Gibbs & Poston, 1975; Becker & Murphy, 1992; Arora et al., 2014). Before we discuss potential drivers of division of labor in teams, it is

[2] For example, different sets of knowledge may be employed to perform a given activity. Similarly, some team members may not perform certain activities even though they possess the required knowledge.

therefore necessary to develop a framework that allows us to think more systematically about the multi-faceted nature of this construct. This conceptual framework highlights a number of useful conceptual building blocks and also points towards ways to operationalize and analyze division of labor empirically.

We start from the premise that the production of a certain output (e.g., a pin or a piece of new knowledge) requires the completion of a number of distinct activities or tasks (Smith, 1776; Becker & Murphy, 1992). We also consider a set of workers who may perform these activities. In this setup, division of labor is absent if each worker performs the full set of activities. Division of labor is used if each worker performs only a subset of activities and it increases with the degree to which individuals specialize in particular activities.

More specifically, consider the stylized example of two teams shown in Figure 1. For each team, the columns refer to a set of different functional activities that need to be performed to produce an output. The rows refer to team members who collaborate in producing the output by performing these activities. The numbers in each of the resulting cells indicates whether a particular member performs a particular activity (1) or not (0).[3] Figure 1 shows that division of labor can be examined using three complementary perspectives: the degree to which a person specializes in one tasks vs. performs a range of tasks (within individual, across activities), the degree to which an activity is performed by one person versus multiple persons (within activity, across individuals), and the degree to which activities tend to be performed together by the same individuals (across individuals and activities).[4]

From the perspective of individual team members, higher division of labor is reflected in a smaller share of activities performed by each person, i.e., members specialize. Less division of labor implies that an individual performs a broader range of activities. Considering the average share of all activities performed by each member, team A employs higher division of labor than team B: The average member performs only 30% of all tasks (an average of 6/5=1.2 activities out of 4 different activities) compared to 70% in team B (14/5=2.8 out of 4 different activities). In addition to differences across teams, however, we also note differences across individuals within a given team. For example, individual 1 in team A is more specialized than individual 5.

Complementing the perspective of individual team members, we can also consider division of labor from the perspective of activities. In particular, Figure 1 shows that higher division of labor reflects that an activity is performed by only a small share of team members, while lower division of labor implies that an activity is performed by a large share (or all) of the members. The activity-level perspective also suggests that team A employs higher division of labor: each activity is performed by an average of only

[3] For the moment, we abstract from the amount of time spent on an activity.
[4] Our distinction between an individual-level and an activity-level perspective should not be confused with the distinction between horizontal and vertical division of labor made in prior literature (where vertical division of labor refers to the division of labor between production activities and managerial activities).

30% of all team members (6/4=1.5 individuals per activity divided by 5 individuals), compared to 70% in team B (14/4=3.5 out of 5 individuals). Again, we can go beyond averages to consider that some activities involve more division of labor than others. For example, in team A, activity 1 is performed by only 20% of all team members, while activity 2 is performed by 40%.

Thus, both perspectives suggest the same level of division of labor when averaging across individuals and activities (e.g., 30% for team A and 70% for team B). However, if there is heterogeneity in division of labor across individuals or across activities, each perspective provides unique insights into the nature of this heterogeneity – the individual perspective allows us to distinguish *specialist* vs. *generalist* members, while the activity-level perspective allows us to distinguish activities that are *distributed* among many members vs. *concentrated* among a few.[5]

Finally, Figure 1 suggests that we can also look across both individuals and activities to examine the relationships among activities, i.e., which activities tend to be performed jointly by the same team member and which ones tend to be performed by different individuals. For example, in team B, individuals who perform activity 1 always perform activity 2 as well, resulting in a positive correlation. In contrast, individuals who perform activity 1 tend not to be involved in activity 3, resulting in a negative correlation between these two activities.

Figure 1: Conceptualizing division of labor

TEAM A

		Activity				Share of
		1	**2**	**3**	**4**	**activities**
Member	**1**	1	0	0	0	25%
	2	0	1	0	0	25%
	3	0	0	0	1	25%
	4	0	0	0	1	25%
	5	0	1	1	0	50%
Share of						**Avg. 30%**
members		20%	40%	20%	40%	**Avg. 30%**

Correlations

	Act.2	Act.3	Act.4
Act.1	-0.41	-0.25	-0.41
Act.2		0.612	-0.67
Act.3			-0.41

TEAM B

		Activity				Share of
		1	**2**	**3**	**4**	**activities**
Member	**1**	1	1	0	1	75%
	2	1	1	0	1	75%
	3	0	0	1	0	25%
	4	1	1	0	1	75%
	5	1	1	1	1	100%
Share of						**Avg. 70%**
members		80%	80%	40%	80%	**Avg. 70%**

Correlations

	Act.2	Act.3	Act.4
Act.1	1	-0.61	1
Act.2		-0.61	1
Act.3			-0.61

[5] The notion of specialists vs. generalists is used widely in the literature and may refer to different aspects of actors. For example, some authors have used the distinction to refer to the breadth of individuals' knowledge (Teodoridis, 2013). Others refer to actors' adaptation to a range of environmental conditions (Van Tienderen, 1991). Our focus is on the range of functional activities performed.

Taken together, the aggregated (average) individual and activity-level indicators inform us about the *level* of division of labor in a given team (i.e., how much is labor divided). A detailed analysis at the level of individual team members and activities, as well as of the correlations among activities informs us about the particular *pattern* of division of labor (i.e., how exactly is labor divided). Based on this conceptual toolkit, we can now discuss in general terms potential drivers of division of labor in sections 2.2. and 2.3. In section 3, we will then consider how these arguments may apply in the particular context of knowledge production in teams.

2.2 Potential drivers of division of labor

Division of labor may result in a number of benefits (Smith, 1776; Becker & Murphy, 1992). First, specializing in order to repeatedly engage in a particular activity may allow individuals to learn and acquire tacit knowledge about how to perform this activity more efficiently ("learning by doing"). Moreover, learning may become increasingly easy as an individual's stock of knowledge in a particular domain increases (Postrel, 2002). To the extent that such learning benefits are present, a worker should focus on one activity rather than splitting her time across multiple different activities, implying a higher degree of specialization and division of labor.

Second, to the extent that an activity requires worker-specific investments such as tools or instruments but also stocks of knowledge, having one person focus on this activity avoids the duplication of investments and allows the existing capital to be utilized more efficiently. Thus, to the extent that workers have to incur fixed costs to perform an activity, the number of individuals working on that activity should be minimized, implying a higher degree of concentration and division of labor.

Potential benefits from specialization and concentration, however, may be offset by increasing costs of coordinating the activities of different team members, and these costs depend on the nature of the interdependencies between activities (Simon, 1962; Becker & Murphy, 1992). As highlighted by organizational theorists, interdependencies between activities can differ in terms of both the frequency and the directionality of work flows (Thompson, 1967; Van de Ven et al., 1976). In the simplest case, activities are independent from each other and their respective outputs are simply pooled in the end to yield the overall output ("pooled interdependence"). In this case, no coordination is required while workers are performing their tasks. In other cases, activities are interdependent in a sequential manner, i.e., the output of one activity serves as the input into another ("sequential interdependence"). In this case, coordination is required only as workers hand off work outputs across stages of the production process. Third, "reciprocal interdependence" describes tasks that depend on each other such that work flows back-and-forth between workers, requiring a significant amount of coordination and mutual adjustment.

Finally, some activities need to be performed simultaneously, without a measurable temporal lapse in the flow of work ("simultaneous interdependence").

Reciprocal and especially simultaneous interdependence between activities imply a high need for coordination between workers performing the respective activities, resulting in high coordination costs. While the organizational literature has focused on the frequency and timing of interactions between workers, we suggest that in the context of knowledge production, coordination costs also depend on how easy it is to communicate and integrate intermediate knowledge outputs between individuals. Intermediate knowledge is more easily transferred if it can be codified at low cost, while communication and integration across different individuals is more difficult if knowledge is tacit (Polanyi, 1966; Osterloh & Frey, 2000). As such, even for a given type of interdependence between the various stages of the research process, the costs of distributing activities across multiple individuals may depend on the tacit versus codified nature of the intermediate knowledge output that is produced. Taken together, the coordination costs between two activities depend on the frequency and timing of interactions as well as the tacit vs. explicit nature of intermediate knowledge outputs. The higher the coordination costs between two activities, the more advantageous it is to have both performed by the same individual rather than different individuals, possibly resulting in a positive correlation observed between the two activities (see Figure 1).

Interdependencies and coordination costs are important not just *across* different activities, but also *within* a given activity. In particular, if the same activity is distributed among multiple team members, each of them performs only a part of this activity and the "sub-activities" or their outputs have to be coordinated and integrated. Similar to our earlier discussion, the resulting coordination costs are likely to depend on the nature of the interdependence between sub-activities and on the tacit vs. explicit nature of the intermediate knowledge outputs. For example, if the results of different sub-activities can simply be pooled to get the desired output, the activity can easily be "decomposed" into different pieces and different individuals can work autonomously and concurrently (Simon, 1962; Von Hippel, 1990; Franzoni & Sauermann, 2014). In contrast, if sub-activities exhibit reciprocal or simultaneous interdependence, distributing them among multiple individuals would result in high coordination costs. As such, a low degree of decomposability suggests that an activity should be performed by few individuals rather than distributed across many, implying a higher degree of concentration and division of labor.

2.3 Division of labor and team size

Division of labor depends not only on characteristics of the activities but also on the total volume of work that needs to be done, i.e., on the size of the project (Smith, 1776; Stigler, 1951; Becker & Murphy, 1992). Assuming that the total volume of work is the sum of the work required for each activity, a larger project is one that requires more work in some or all of these activities. Assuming further that all individuals work full-time, larger projects require larger teams.[6] We now turn to the question how division of labor may differ in teams of different size.

For very small projects, opportunities for division of labor are limited because there is not enough volume of work in each activity to allow one person to work on this activity full time (Smith, 1776). To illustrate, consider team C in Figure 2. This team's project requires the total efforts of only two people, but still involves all 4 activities. As such, team members have to perform multiple activities. If the total volume of work is larger, however, the team can have more members and there may be enough work in a given activity to have one team member specialize in that activity (Team D). As such, if there are benefits from specialization, division of labor will be higher in a larger team. Of course, our earlier discussion of coordination costs across activities still applies, and team members may not specialize fully.

As project size increases further, teams will be larger but the potential for division of labor across activities is exhausted (i.e., team members are already specialists as much as this is beneficial). Any addition of team members has the primary benefit of allowing the parallel work of multiple individuals in a given activity. Consider team E in Figure 2. This team has the same level of division of labor as Team D but each activity is performed by twice the number of individuals, allowing a larger volume of work to be done in a given amount of time.

While additional members may in principle be distributed evenly across activities, the degree to which teams take advantage of parallel work within a given activity may differ across activities for two reasons. First, some activities may require only a relatively fixed volume of work regardless of the size of the overall project, while the volume of work required for other activities may vary more across projects. Second, even if the volume of work in an activity is larger, it may be difficult to distribute the activity among multiple team members if the activity's decomposability is low.[7] For either or both of these reasons, the number of individuals engaged in a particular activity may grow slower with team size than for other activities. Compare teams D and F in Figure 2. Team F employs twice the number of individuals as team D, but the additional team members are all engaged in activities 2 and 4, while the number of individuals engaged in activities 1 and 3 remains the same as in team D. As such, the share of members engaged in activities 1 and 3 is lower, while that engaged in activities 2 and 4 is higher than in team D.

Overall, larger team size will be associated with higher division of labor up to a certain point, after which additional members are likely to increase parallel processing within a given activity. While the share of members engaged in each activity may remain stable in principle, activities that exhibit

---

[6] Teams could stay small and increase the overall duration of the project. Given the importance of speed in scientific competition (Freeman et al., 2001) we assume that increasing team size is the primary mechanism to deal with bigger problems.
[7] As such, the only way to get this work done is to have a small number of people work for a longer duration each.

higher cross-project variation in the volume of work or that are more decomposable may employ a larger share of team members in larger teams.

Figure 2: Division of labor and project size

TEAM C

		Activity 1	2	3	4	Share of activities
Member	**1**	1	1	0	0	50%
	2	0	0	1	1	50%
Share of						**Avg. 50%**
members		50%	50%	50%	50%	**Avg. 50%**

Correlations

	Act.2	Act.3	Act.4
Act.1	1	-1	-1
Act.2		-1	-1
Act.3			1

TEAM D

		Activity 1	2	3	4	Share of activities
Member	**1**	1	0	0	0	25%
	2	0	1	0	0	25%
	3	0	0	1	0	25%
	4	0	0	0	1	25%
Share of						**Avg. 25%**
members		25%	25%	25%	25%	**Avg. 25%**

Correlations

	Act.2	Act.3	Act.4
Act.1	-0.33	-0.33	-0.33
Act.2		-0.33	-0.33
Act.3			-0.33

TEAM E

		Activity 1	2	3	4	Share of activities
Member	**1**	1	0	0	0	25%
	2	0	1	0	0	25%
	3	0	0	1	0	25%
	4	0	0	0	1	25%
	5	1	0	0	0	25%
	6	0	1	0	0	25%
	7	0	0	1	0	25%
	8	0	0	0	1	25%
Share of						**Avg. 25%**
members		25%	25%	25%	25%	**Avg. 25%**

Correlations

	Act.2	Act.3	Act.4
Act.1	-0.33	-0.33	-0.33
Act.2		-0.33	-0.33
Act.3			-0.33

TEAM F

		Activity 1	2	3	4	Share of activities
Member	**1**	1	0	0	0	25%
	2	0	1	0	0	25%
	3	0	0	1	0	25%
	4	0	0	0	1	25%
	5	0	1	0	0	25%
	6	0	1	0	0	25%
	7	0	0	0	1	25%
	8	0	0	0	1	25%
Share of						**Avg. 25%**
members		13%	38%	13%	38%	**Avg. 25%**

Correlations

	Act.2	Act.3	Act.4
Act.1	-0.29	-0.14	-0.29
Act.2		-0.29	-0.6
Act.3			-0.29

3 Applying the Framework to Scientific Teams

We now turn from the general discussion of division of labor in teams to consider key issues more concretely in the context of scientific teams.

3.1 Types of activities and potential benefits from specialization

We first characterize in a stylized way five activities that are frequently required in knowledge production, especially in the empirical sciences. While different classifications of activities could be used, our distinction is consistent with schemes commonly used in the scholarly literature, by editors, as well as scientists themselves (ICMJE, 2010; Venkatraman, 2010; Marušić et al., 2011; Shibayama et al., 2013; Allen et al., 2014).[8] For each activity, we also consider potential benefits from specialization within a given research project.[9]

Conceptualizing a research study involves defining research questions, formulating hypotheses, and designing experiments to test the hypotheses. Identifying an important research problem is often seen as the key challenge in research and the key predictor of impact in the broader community (Merton, 1973; Laudel, 2002). As such, conceptual activities require deep knowledge of the research field as well as the ability to systematically build theoretical arguments (Merton, 1973; Delamont & Atkinson, 2001). While individuals may acquire such knowledge and improve their conceptual skills over the careers, learning within a given project is limited since these activities are not repeated frequently, suggesting that learning benefits from specialization within a given team are likely to be low. Conceptual activities also do not require investments in physical infrastructure, suggesting no efficiency gains from specialization in that respect.

Obtaining data. Most research studies utilize data to test hypotheses or examine empirical regularities. Data can be obtained by running experiments or by observing and measuring objects in their natural environments. Activities related to running experiments or collecting observational data may exhibit strong benefits from specialization because they often rely on costly physical infrastructure such as sequencers or telescopes (Stephan, 2012) and also typically benefit from learning and experience with a given procedure (Laudel, 2002). For example, Barley argues that individuals performing empirical tasks often acquire semiotic knowledge that allows them to make sense of subtle differences when peering through the microscope and to have a feeling for observational data collection, materials, instruments and techniques (Barley, 1996).

Analyzing data. Once data have been obtained, significant effort is involved in analyzing these data using a variety of statistical methods. Data analysis may involve significant benefits from specialization since it often involves learning about the intricacies of data sets as well as the utilization of specialized hardware and statistical methods.

Writing a research paper involves the integration of conceptual ideas and empirical results into a document that can be shared with the broader research community. This process involves placing the work in a particular stream of research, describing the gap in prior literature, conveying the logic of the theoretical arguments, describing empirical results, and discussing limitations and implications for future

[8] Of course, this classification is very coarse. Not all activities are involved in all projects, and the relative importance of these activities may differ across fields. Moreover, while we discuss potential benefits from specialization and interdependencies between activities in qualitative terms, their particular nature may differ across fields as well (Shibayama et al., 2013). We will explore differences in division of labor across fields empirically in section 5.3.

[9] Throughout this paper, we focus on a given project and, as such, abstract from potential benefits when a scientist specializes in particular activities across multiple projects.

research. As such, writing requires both intellectual and social skills (Shibayama et al., 2013; Fayard & Metiu, forthcoming). While writing is less repetitive than empirical activities, it is likely more so than conceptual activities discussed above, and there may be learning benefits in a given project. At the same time, little physical infrastructure is required, suggesting that overall the benefits from specialization in writing may be limited.

3.2 Division of labor and team size

Our theoretical discussion suggested that the number of individuals engaged in particular activities may expand at different rates with team size because of differences in the volume of work or in the degree of decomposability. In the context of knowledge production, we suggest that the required volume of work in writing and conceptualization varies less across projects than the work required to collect and analyze data. In other words, larger projects are larger primarily because they require more empirical work, not because they require more conceptualization or writing. As such, larger teams should need additional team members primarily for data collection and analysis. As per our discussion above, we also expect that decomposability is higher for empirical activities than for conceptualization and writing. Taken together, these arguments suggest that as team size increases, the number of workers engaged in data collection and analysis tasks is likely to grow faster than the number of people engaged in conceptualization and writing.

3.3 Summary

Our discussion resulted in a number of general predictions:

1. Distributed vs. concentrated activities: An activity will be performed by a larger share of team members (be more distributed) if it involves lower benefits from specialization, a larger volume of work, and a higher degree of decomposability.
2. Correlations among activities: Activities will be performed jointly (by less specialized team members) if they involve low benefits from specialization and high interdependencies with each other.
3. Division of labor and team size: Larger teams have more specialized team members and activities tend to be concentrated among a smaller share of members. Concentration increases more rapidly with team size for activities that have less variation in volume across projects and that have low decomposability.

Table 1 summarizes our priors regarding potential drivers of division of labor in the context of scientific teams. Given the large number of factors, potentially offsetting effects, and the largely qualitative nature of our priors, we refrain from making specific predictions. However, Table 1 will be useful in interpreting the results of the following empirical analysis.

Table 1: Conjectures regarding drivers of division of labor in the context of scientific research

	Benefits from specialization across activities	Interdependencies across activities			Variation in volume of work across projects	Decompos-ability
		Performing/ Collecting	Analyzing	Writing		
Conceptualizing	Low	Sequential	Sequential	Reciprocal/ Simultaneous	Low	Low
Performing experiment, collecting observational data	High		Reciprocal	Sequential	High	High
Analyzing data	High			Reciprocal/ Simultaneous	High	Medium
Write	Medium				Low	Low

4 Data and Measurement

4.1 Data

Our empirical analysis draws on data from articles published in the journal PLOS ONE. This journal was co-founded in 2006 by a Nobel Prize winner and a former director of NIH and is the largest Open Access peer-reviewed journal.[10] PLOS ONE covers primary research from any discipline within science and medicine and is ranked in the top quartile in the field of interdisciplinary sciences by ISI Web of Science, with an impact factor of around 4. PLOS ONE has a strong reputation in the scientific community and a recent study published in *Science* highlights particularly the journal's rigorous and transparent peer review process (Bohannon, 2013). While our featured analysis exploits the full breadth of data, we also perform auxiliary analyses using only the top 10 percent of papers in terms of their citation impact.

Data from PLOS ONE are particularly appealing for two reasons. First, they allow us to gain insights into a broad range of scientific projects since articles vary considerably in terms of both fields and article impact. Second, PLOS ONE is a pioneer in requiring authors to disclose the contributions made by each team member during the submission process. While contributions are still reported by the authors, they are likely to reflect an agreed-upon assessment of contributions by the full team of authors

[10] http://www.plos.org/about/; http://richardpoynder.co.uk/PLoS_ONE.pdf

and, thus, are less affected by biases than surveyed self-reports of individual authors (Haeussler & Sauermann, 2013; Shibayama et al., 2013).

We use information from 13,935 articles published from February 2007 to September 2011.[11] These articles listed 91,311 authors. We include only original research articles; editorials, errata, and letters are excluded. Since we are interested in the division of labor within teams, we dropped 233 single authored papers, 169 papers that did not disclose the contribution of one or more authors, 16 papers listing only "other contributions", and 61 papers that did not list any authors as having "written" (since each paper needs to be written, the contribution statements of these papers are likely incomplete). For the main analysis, we exclude papers with more than 20 authors (N=150) since the organization of knowledge production may be qualitatively different in big science projects and since the small cell sizes make team-level analyses of large teams difficult. We complement the information obtained from the articles with information on the citations received by each article, provided by PLOS ONE. In addition, we collected data on the quantity and quality of authors' prior publications using the Scopus database.

4.2 Measures

Table 2 shows summary statistics for individual level variables as well as for team level variables.

4.2.1 Key Measures

Types of contributions. Upon submission of a manuscript, PLOS ONE asks authors to specify the particular contributions made by each individual author. To do so, the journal offers a form with five pre-defined types of contributions: (1) conceived and designed the study (*conceived*) (2) performed the experiments (*performed*), (3) contributed materials, such as physical inputs/reagents/analysis tools (*materials*), (4) analyzed the data (*analyzed*), and (5) wrote the paper (*wrote*). These activities map quite well to the activities we discussed in section 3, except that we did not discuss contributions of materials, which tend to be less common (see below).[12] For these 5 standard contributions, authors enter the respective co-authors' initials into the corresponding box. For non-standard contributions, a free-text field "other" allows authors to indicate additional contributions and which individuals have made them. We create a dummy variable indicating whether a particular individual was listed as having made some other contribution (*othercontribution*).

We recognize that listed contributions may be imperfect measures of the activities performed by the individuals who participated in a project. In particular, our measure does not capture work done by "ghost authors" who made significant contributions but are not listed as authors on the paper, although the

[11] We excluded the first 200 papers published in PLOS ONE to account for the community having to acquaint itself to the new journal and its processes. Including these papers does not change our results.
[12] http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

share of such individuals in teams tends to be very small (Haeussler & Sauermann, 2013; Lissoni et al., 2013). Relatedly, the listing of author contributions may reflect not only objective contributions but also a social process of negotiation among team members, with more powerful team members potentially negotiating to be listed as having made contributions they did not actually make. We will address the latter possibility by including a number of individual-level control variables (see below). Finally, the contribution measures are dichotomous, informing us that team members have contributed above a certain minimum threshold to be listed under a particular activity, but providing little insight into the extent of their contributions beyond that threshold. Despite its limitations, our measure has key advantages over available alternatives. Most obviously, the measures allow us to examine division of labor in a very large sample of projects, complementing prior qualitative work using small numbers of cases (Owen-Smith, 2001). By using pre-defined categories, we obtain measures that are easily compared across teams, while relying on the scientists themselves (rather than less knowledgeable coders) to decide which categories best fit the contributions made by the various team members. Finally, while information about author contributions can also be obtained through surveys, individuals may overestimate their contributions to a team effort (Kamo, 2000; Ivaniš et al., 2011). The contributions listed on PLOS papers should be less affected by such biases to the extent that they reflect a consensus assessment of team members, all of whom realize that giving "too much" credit to another team member may reduce the credit allocated to them.

Division of labor. Mirroring our conceptual discussion, we capture division of labor using a number of complementary measures.

First, we count the number of activities performed by a given author and divide it by the total number of types of contributions listed on the paper (*shareofcontributions*). This ratio reflects what share of the up to 6 listed types of contributions a particular author performed. We aggregate this measure to the team level by computing the average share of contributions across all team members, resulting in *t_avg_shareofcontributions*. This variable equals 1 in the extreme case that all individuals perform all the activities that are listed on the paper (i.e., using no specialization). The measure is lowest (with a theoretical minimum of 0.167) for teams where each individual performs only one of six listed activities. To explore heterogeneity in the degree of team members' specialization, we also compute the standard deviation of *shareofcontributions* within a given team (*t_stdev_shareofcontributions*).

Finally, we obtain an activity-level measure of division of labor by counting the number of authors that performed a given activity and dividing this count by the total number of authors on the paper. The resulting 6 measures capture what share of all team members has performed a particular activity (*t_shareconceived*, *t_shareperformed*, *t_sharematerials*, *t_shareanalyzed*, *t_sharewrote*, and *t_shareother*). Note that these measures are coded as missing (rather than 0) if a particular activity is not listed on the paper at all.

Team Size. *T_teamsize* is the count of the number of co-authors on the paper.

4.2.2 Control variables

We include a number of variables to control for heterogeneity at the level of the project/team as well as the level of individual authors.

Project/team level control variables. To control for differences in the nature of research and of resource availability, we include the indicator variable *t_funding_basic* that equals one if the paper acknowledges funding from a governmental basic research funding agency such as the National Science Foundation, National Institute of Health, German Research Foundation, or the World Health Organization. The variable *t_funding_industry* equals one if the paper acknowledges funding from a for-profit organization. The variable *t_funding_other* equals one if a paper acknowledges funding from another source. We further control for differences in the nature of research by including fixed effects for 42 *fields* (+1 "other") based on field classifications listed on the article. Note that each paper can be classified into multiple fields, and the average number of field classifications is 2.49.

We also include a number of variables to control for differences in the composition and location of teams. The measure *t_affiliations* counts the number of unique author affiliations listed on the paper. The dummy *t_country_US* equals 1 if the first author's main affiliation is in the US, and *t_country_developing* equals 1 if the first author's main affiliation is in a developing country (other countries is the omitted category).[13] We code the variable *t_firm* as 1 if at least one of the author affiliations is with a firm.[14] Finally, we include as additional article level controls the number of pages (*t_pages)*, the publication date (*t_publication date*), and the number of listed research activities (up to 6, *t_totalactivitieslisted*).

Individual level control variables. We code a number of variables based on the affiliations listed for each author. We measure the reputation of an author's listed institution by the variable *institutionrank* which we coded according to the 2012 Academic Ranking of World Top 500 Universities (the so-called Shanghai ranking). The variable is 1 if an author's institution belongs to the Top 20 universities, 2 for Top 21-50, 3 for Top 51-100, 4 for Top 101-200, 5 for Top 201-300, 6 for Top 301-

[13] We use the affiliation of the first author since this person likely contributed most to the project and his/her first listed affiliation is clearly his/her main affiliation. For other authors, the first listed affiliation does not have to be the main affiliation since PLOS ONE uses one set of affiliation identifiers in running order. As such, an affiliation that appears for an author early in the author list will appear first for a later author even if it is not the later author's main affiliation.
[14] We searched for terms such as "corporation", "inc", and "S.A.". We also manually checked author affiliations to identify commonly occurring firm names such as Pfizer or Roche.

400, 7 for Top 401-500 and 8 for non-listed universities. The indicator variable *developing country* is one if an author has at least one affiliation in a developing country and zero otherwise.

In addition to information taken from the PLOS articles, we obtained data on authors' prior publications using the Scopus database. We use the number of publications over the five years prior to the focal PLOS article as our measure of the quantity of prior publications (*priorpubs_quantity*). We use the average yearly number of citations to these articles (dividing total yearly citation counts by the total number of publications) as a proxy for the quality of prior publications (*priorpubs_quality*).[15] We estimate *professionalage* as the difference between the year of the PLOS paper and the year of the first publication. To reduce the incidence of erroneous matching of author names, we manually checked all cases in which the search algorithm returned more than 200 publications for the time period of five years prior to the focal PLOS publication.[16] We also recoded Scopus measures for a small number of individuals for whom the search algorithm returned publications that were more than 60 years old. If we were not able to reliably identify authors in Scopus, we code the main Scopus variables as 0 and additional code the variable *missingscopus* as 1 (and zero otherwise).

Including these measures of authors' prior publications and professional age is useful to address the concern that the division of labor may not simply reflect efficiency considerations but also that more senior or more accomplished team members are able to choose to work on activities they prefer. Note that we are examining overall patterns of division of labor in teams rather than the question which particular individuals perform which (sets of) activities. As such, any such social mechanisms would be of interest only if they led teams to divide activities in ways that are less optimal from an efficiency perspective. Although we explicit control for measures of social status, this potential mechanism should be kept in mind when interpreting the empirical results.

--- Table 2 here ---

5 Results

5.1 Patterns of division of labor within individuals and activities

The summary statistics in Table 2 shows that the average project lists 4.8 out of 6 possible activities. Figure 3, panel A shows the share of all papers that list a particular activity and suggests that conception, performing the experiment, analysis, and writing are part of virtually every paper in the

[15] For authors without matching Scopus data, these variables are set to zero.

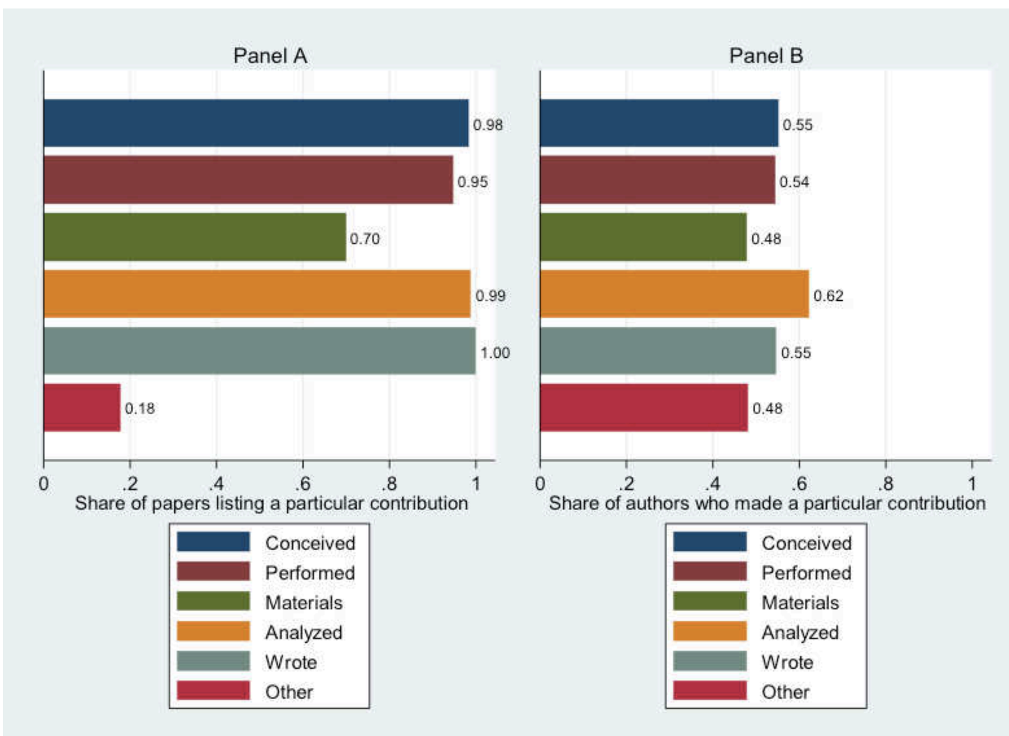
[16] When matching seemed erroneous, we replaced Scopus variables as missing. For example, our tool returned 1101 publications for the five years prior to the focal PLOS publication for the very common name *Ying LIU* and 783 publications for *Lin XU*. On the other hand, our tool correctly identified 404 publications for the French researcher *Didier RAOULT*, who is known to have the highest number of publications to date in France (see http://en.wikipedia.org/wiki/Didier_Raoult).

sample. The share of papers that list contributions of materials/data is somewhat lower (70%). Only 18% of papers list some "other" contribution, suggesting that the 5 pre-defined contributions are sufficient to describe the roles of most authors.[17]

To give a first impression of division of labor from the perspective of activities, Panel B in Figure 3 shows the share of authors who is listed as having made a particular contribution, conditional upon that activity being listed at all. Across all teams, the shares range from 48% (share of authors who made an "other" contribution if such a contribution was listed) to 62% (share of authors who analyzed data).

At the level of the individual, we find that the average author performed 51% of all listed activities on a paper (*shareofcontributions*), for an average of 2.46 different contributions (Table 2). A look at the distribution of this measure, however, reveals considerable heterogeneity: Roughly 22% of team members have a *shareofcontributions* of 0.2 or less, indicating a very high degree of specialization. At the other extreme, 11% of team members have a share of contributions of greater than 0.8 and 17% a share between 0.6 and 0.8, making them "generalists" in terms of the range of activities performed.

Figure 3: Share of papers listing a particular contribution and share of authors performing each

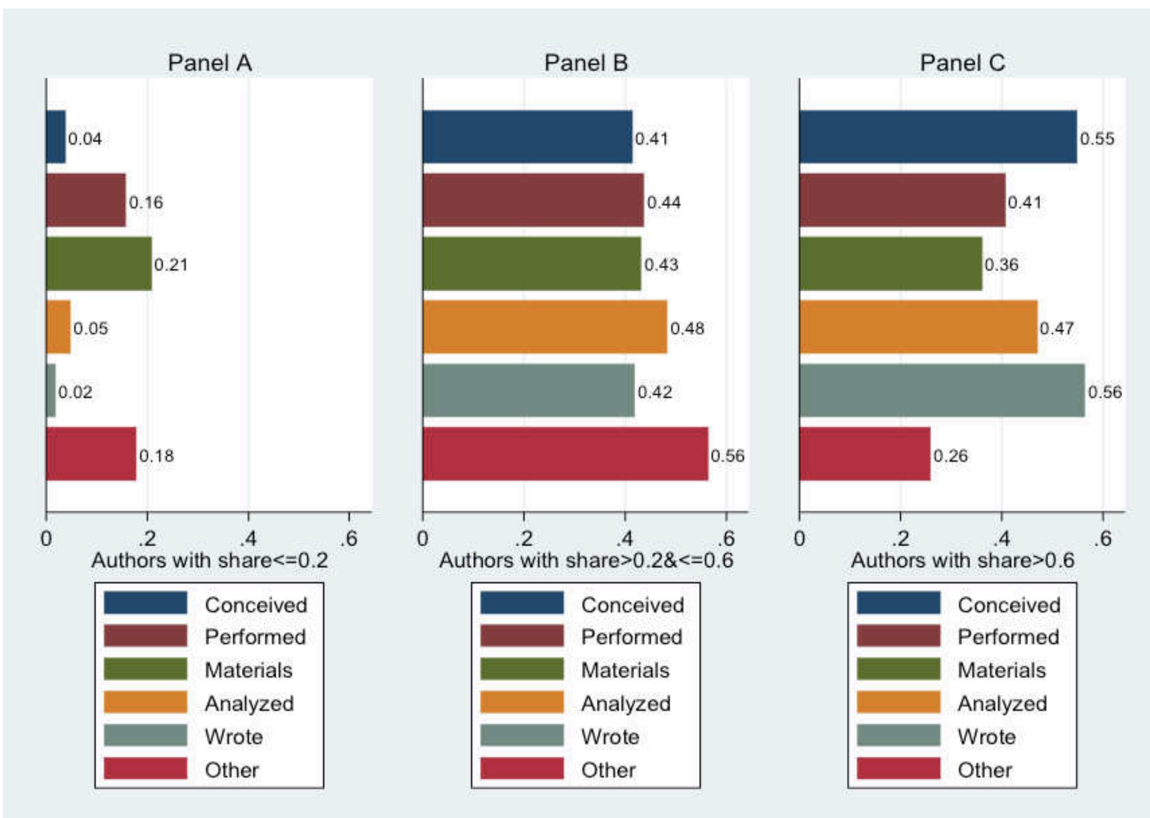


Note: Percentages in panel B are computed conditional upon an activity being listed on the paper.

[17] For example, the "other" contributions included items such as "developed design of simulation", "critically revised the method" or "performed subject recruitment".

Given the considerable heterogeneity in team members' levels of specialization, we can look across both individuals and activities to examine whether certain activities tend to be performed by specialized individuals while others tend to be performed by generalists. For that purpose, we focus on the authors listed as having performed a particular activity and compute the share of these authors with high specialization (*shareofcontributions*<=0.2, Figure 4, panel A), medium specialization (*shareofcontributions*>0.2 but <=0.6, panel B) and low specialization (*shareofcontributions*>0.6, panel C). We see that *performed* and *materials* (as well as *other*) have the highest share of highly specialized team members and the lowest share of generalist team members. In contrast, *conceived* and *wrote* have the smallest shares of specialists and the largest share of generalists, consistent with our conjectures that these particular activities benefit less from specialization and that they may exhibit important interdependencies with other activities.

Figure 4: Share of team members who performed a particular activity and have a high level of specialization (panel A), medium specialization (panel B), low specialization (panel C)



To further probe the relationships among activities, we perform a series of regression analyses. In particular, we regress a focal activity on the set of the 5 remaining ones.[18] Since we are interested in how activities are correlated within a given team, we estimate these models using linear probability models (LPM) with paper fixed effects. The coefficients of LPMs can be interpreted as differences in probabilities and the paper fixed effects control for unobserved heterogeneity at the level of the project (e.g., the nature of research). In addition, these regressions include a number of individual level controls such as measures of team members' prior performance, professional age, and institutional background.

Model 1 in Table 3 uses as dependent variable the dummy indicating whether a team member has contributed by writing the paper (*wrote*). As expected, we find a strong positive relationship with *analyzed* and *conceived*, consistent with the notion that writing involves the integration of conceptual arguments and empirical results and that there are interdependencies between these three activities that make it beneficial for the same team members to be engaged in all three. Model 2 uses as dependent variable whether an individual has *analyzed*. The strongest relationship emerges again with *wrote*. Individuals who *performed* the experiment are more likely to also be involved in data analysis, consistent with the notion that generating data and analyzing data show strong interdependencies. In contrast, individuals who contributed *materials* are less likely to also have *analyzed*, consistent with the notion that the provision of materials and the analysis of data are largely sequential in nature. Model 3 focuses on the predictors of *materials* and shows that it has a negative relationship with all other activities. The negative coefficient is largest for *performed*, perhaps reflecting that performing experiments contributing materials (including data) are alternative ways to obtain data and that a given individual will do either one but not both. Model 4 reinforces this finding by showing a strong negative relationship between *performed* and *materials*, while also showing a significant positive relationship between *performed* and *analyzed*. Finally, Model 5 examines which activities predict that a team member has *conceived*. A notable observation is that *wrote* has a strong positive coefficient, and this coefficient is larger than the coefficient of *conceived* in the regression of *wrote* (0.490 vs. 0.424 in model 1). This finding might suggest an asymmetric relationship such that (good) writing is more dependent on being involved in conceptual activities than the other way around.

Overall, these regressions suggest systematic patterns in the relationships among activities. Individuals who write the paper tend to also be engaged in analysis and conception, while individuals who perform the experiment are likely to also be involved in data analysis. The provision of materials tends to be relatively independent or even isolated.

[18] An alternative approach would be to perform a factor analysis. Using this approach, one factor with eigenvalue >1 emerges, with the largest loadings for *conceived*/*wrote*/*analyzed*. Two smaller factors are below the eigenvalue>1 threshold, one combining *performed*/*analyzed*, and the other including only *materials*. While these results are consistent with our regression approach, the latter provides more detailed insights into the relationships between each pair of activities.

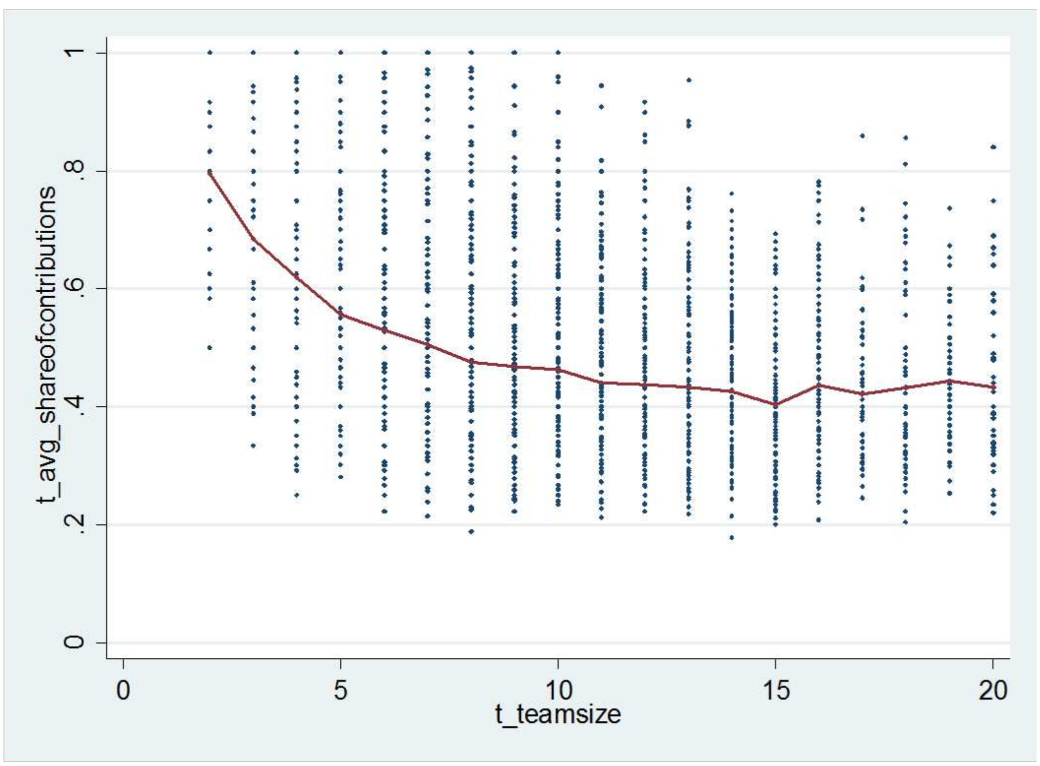
--- Table 3 here ---

5.2 Division of labor and team size

5.2.1 Descriptive results

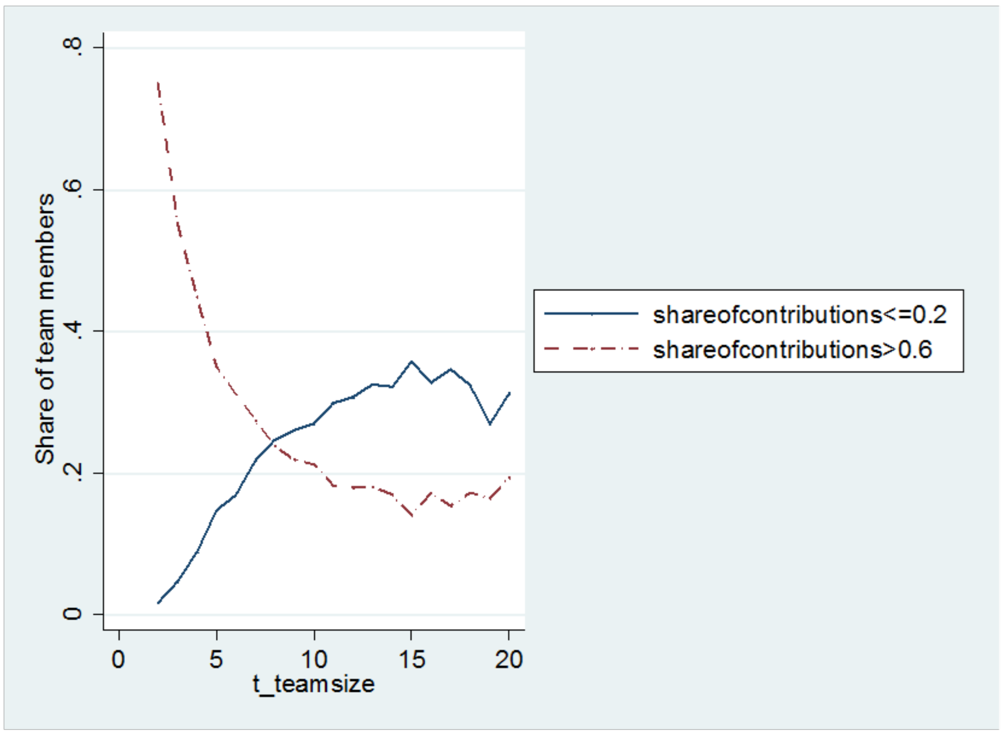
We now examine how division of labor differs across teams of different size. Figure 5 plots *t_avg_shareofcontributions* against the size of the team, allowing us to examine the relationship between team size and team members' specialization. Figure 5 shows a negative relationship: while the average member in teams of 2 performs roughly 80% of all listed contributions, that share decreases to around 43% of listed contributions in teams of 14 or more. Recall that the minimum of this measure is roughly 0.2 (0.163 in projects with 6 activities and 0.2 in projects with 5 activities), yet the average share of listed contributions stabilizes above 0.4, suggesting that perfect specialization of team members is not advantageous.

Figure 5: Average share of listed activities performed by a team member, by team size



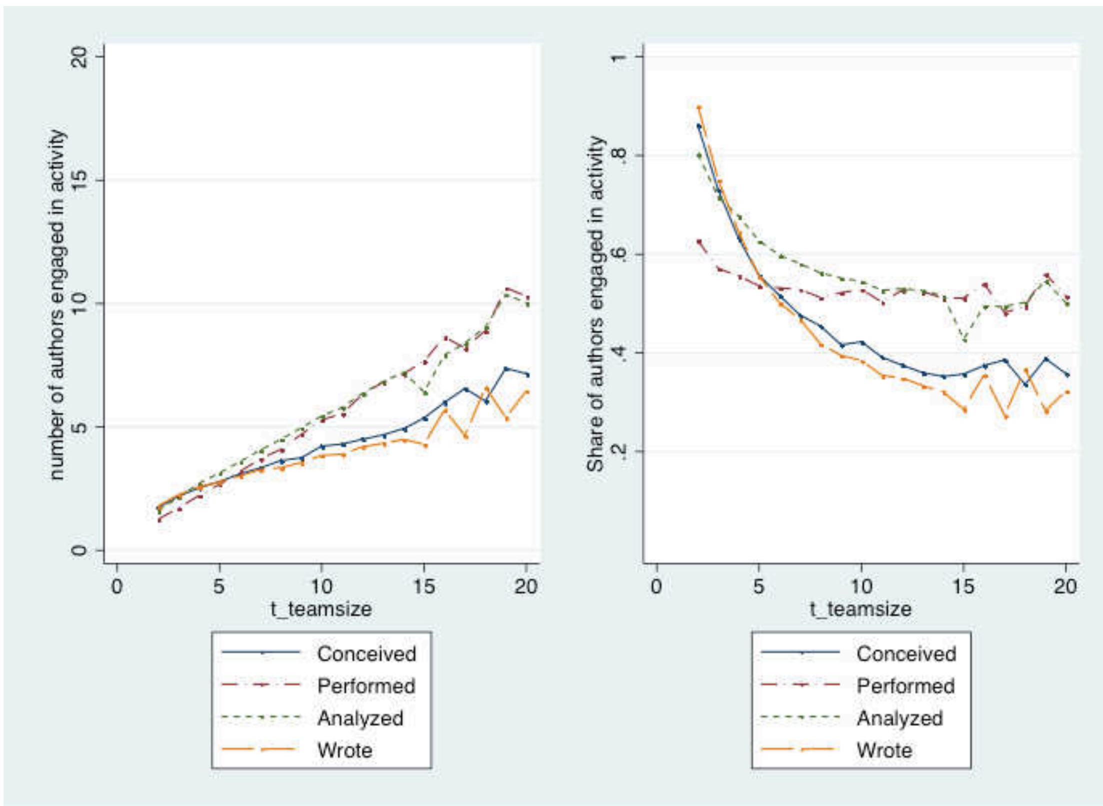
To further probe the limits to specialization, we examine the shares of specialists (*shareofcontributions*<=0.2) and generalists (*shareofcontributions* between 0.6 and 1) for each size class (Figure 6). We observe that the share of specialists increase sharply with team size but then stabilizes at a team size of roughly 15, reaching around 30%. The share of generalists decreases until about a teamsize of 10 but then remains quite stable at roughly 18%. Thus, while larger teams employ higher division of labor, most individuals do not specialize perfectly. Moreover, even large teams have generalists who perform a broad range of activities, perhaps because such individuals help teams in overcoming the coordination and integration challenges associated with higher division of labor.

Figure 6: Share of specialists and generalists, by team size



Figures 5 and 6 above showed an increasing specialization at the individual level, and we expect a corresponding increase in the concentration of activities (i.e., activities are performed by smaller shares of team members). However, we conjectured that changes in division of labor at the activity level may differ depending on the characteristics of the activity. Panel A in Figure 7 plots the number of team members who perform a given activity by team size, and Panel B shows the corresponding share of team members (conditional upon an activity being listed on the paper). As expected, the raw number of individuals performing a given activity increases with team size. Consistent with a general increase in division of labor, however, the share of individuals engaged in a particular activity decreases for all activities, i.e., activities become more concentrated. Most importantly, we find as predicted that concentration increases less for *performed* and *analyzed* than for *conceived* and *wrote*. For example, the share of team members involved in writing drops from 90% in teams of 2 to 32% in teams of 20. The share of team members who *performed* decreases less strongly from 63% to 51%. This leads to an interesting inversion: writing and conceptualization tend to be more distributed than empirical activities in small teams, but empirical activities tend to be more distributed than writing and conceptualization in large teams.

Figure 7: Number and share of authors engaged in particular activities, by team size



5.2.2 Regressions

We now examine the relationship between team size and division of labor using regression analysis, which allows us to control for a number of potentially confounding variables (Table 4). In model 1, we regress *t_avg_shareofcontributions* on dummy variables for each team size category and project-level control variables. Consistent with Figure 5 above, we see significant negative coefficients of team size that stabilize at a team size of around 15. Complementing this analysis, model 2 explores the relationship between team size and *t_stdev_shareofcontributions*. This measure initially increases with team size, i.e., there is a greater variation in the degree to which some individuals specialize in a small number of activities, while others perform a broader range of activities. However, the coefficients decrease again for larger team sizes and *t_stdev_shareofcontributions* in teams above 15 is not significantly different from that in teams with only 2 members. The explanation becomes clear in Figure 6 above. In particular, the standard deviation of *shareofcontributions* should be high if teams include substantial shares of both specialists and generalists, but it will be low if teams are composed primarily of one group. Given the declining share of generalists and increasing share of specialists as team size grows, it is the medium size teams that have the most diverse mix of members in terms of their levels of specialization.

Models 3-8 regress the share of individuals engaged in each of the 6 activities on team size, contingent upon a project listing the particular activity at all. Consistent with Figure 7, we see negative coefficients for the team size dummies, indicating increasing concentration. However, this relationship is much stronger for *conceived* and *wrote* than for the other activities. Indeed, controlling for project characteristics, *t_shareperformed* is virtually stable across the team size distribution, suggesting that the number of team members who perform the experiment increases linearly with team size.

--- Table 4 here ---

5.3 Differences in division of labor across types of projects and institutional contexts

The prior sections examined levels and patterns of division of labor across a broad sample of teams. We now explore potential differences in the division of labor across different types of projects or institutional contexts.

In a first set of regressions in Table 5, we explore differences across scientific fields. For that purpose, we aggregate the large number of detailed field dummies into broader main fields, including biology (84%), medicine/health (52%), physical sciences (4%), earth sciences (2%), computer/information sciences (1%), and other (<1%). Recall that each paper can be categorized into multiple fields. Model 1 uses *t_avg_shareofcontributions* as a measure of division of labor at the team level, where a higher score indicates that team members tend to be less specialized and thus use less division of labor. We find that papers that are categorized in biology (vs. not categorized in biology) use more division of labor, while papers in the earth sciences use less division of labor. Controlling for other field categorizations, papers in medicine/health also use less division of labor. Models 2-7 focus on the share of team members performing particular activities. We find that teams in biology have a smaller share of individuals listed as having conceived or written, but higher shares of individuals listed as having analyzed data. Similarly, papers classified in medicine/health also have a higher share of individuals involved in data analysis. In contrast, papers in the physical sciences have a higher share of team members who wrote or contributed materials, but fewer who performed experiments. While we had no priors regarding these field differences, they provide interesting initial insights into differences in the way knowledge is produced across fields. Note that these differences may reflect both differences in the volume of work required for certain activities, but also differences in the way in which teams divide labor.

Second, we explore whether division of labor differs depending on the degree of interdisciplinarity of a research project. To proxy for interdisciplinarity, we count the number of different (detailed) field classifications listed on the paper and then form three categories indicating whether a paper was classified in only one field (19.63%), two or three fields (64.85%), or more than 3 fields (15.51%). Model 8 shows that division of labor is lower in projects that list more than 3 different field classifications, and models 9-14 suggest that this reflects a larger share of team members being involved in performing experiments and analyzing data, without a decrease in the share of individuals engaged in other activities. One potential interpretation is that interdisciplinary projects draw on a broader set of empirical techniques, thus limiting teams' ability to concentrate empirical work among a smaller set of specialized team members.

Third, we conjectured that division of labor may be different in novel fields than in fields that are more established. To obtain a proxy for the age of a field, we identified for each field the main professional society or association and determined the year of its founding. We then classified fields as "new" if the society/association was founded after 1970. We then classified papers according to the youngest listed field, resulting in 20.86% of papers classified as in a new field and 79.14% as in an established field (*t_field_new*). We find that division of labor is higher in projects involving novel fields (model 8, which reflects that a smaller share of members is involved in conception and, especially, performing the experiment (models 9-14). One potential explanation is that newer fields require larger investments into learning new empirical methods. As per our theory, such fixed costs would lead to benefits from specialization, leading teams to allocate these activities to a smaller share of their members.

In a final analysis we explore whether there are differences in the division of labor between teams in different institutional spheres, i.e., academic versus industrial science. On the one hand, the efficient production of publications should be an important consideration in both sectors, and the notion of the "lab as a firm" has been raised particularly in the context of the highly competitive academic sciences (Hackett, 1990; Freeman et al., 2001; Stephan, 2012). On the other hand, academic labs may also pursue other goals, including the training of junior scientists. If such training goals include the exposure to different types of research activities, academic labs may employ less division of labor than industrial labs (Shibayama et al., 2013). To explore this possibility, we inspect the coefficient of *t_firm*, a variable that takes on the value of 1 if at least one team member is affiliated with a firm. We find no significant difference in the average share of contributions performed by team members (model 8). Models 12 and 13 show a slightly higher share of team members who *analyzed* and a smaller share who *wrote* in firms, but these differences are relatively small and significant only at the 5% level. Thus, we find no qualitative differences in the division of labor between teams operating in the two institutional spheres.

--- Table 5 here ---

6 Conclusion and Future Research

An increasing amount of knowledge is created collaboratively in teams (Wuchty et al., 2007; Stephan, 2012), yet little is known about how work is divided and distributed among team members. Conceptualizing research as a process that involves a number of distinct activities such as conceptualization, data collection, and analysis, we developed a conceptual framework to analyze division of labor in teams using multiple perspectives. Focusing on individual team members, the first perspective provides insight into authors' levels of specialization. Focusing on particular activities, the second perspective provides insights into the degree to which activities are distributed across multiple team members vs. concentrated among a few. Considering individuals and activities jointly, the third perspective provides insights into which activities tend to be performed together by the same individuals, and which activities tend to be performed by more highly specialized team members. We draw on prior organizational literature to discuss a number of potential drivers of division of labor, including benefits from specialization, interdependencies among activities, and the decomposability of activities. We also consider how division of labor relates to the size of the project.

Drawing on this framework, we then provided empirical evidence on division of labor in scientific teams based on data from over 13,000 scientific articles disclosing the contributions of individual authors. In addition to providing nuanced descriptive results that should be of interest in themselves, our empirical analysis results in a number of more general insights. First, while division of labor can be characterized using summary measures at the level of the team, there is also important and systematic heterogeneity across individual team members and activities that deserves our attention. Second, the higher level of granularity provided by author and activity level measures allows us to draw on a rich set of mechanisms and theories developed in prior literature to consider potential drivers of the observed patterns of division of labor. Taken together, our conceptual framework and empirical insights contributes to the literature on collective knowledge production, organizational theory and economics of science literature as well as provides a linkage to connect these literature streams.

Our study adds to the literature on knowledge production in teams and the economics of science. While knowledge production is often viewed as the recombination of existing knowledge (Simonton, 2003; Singh & Fleming, 2010; Gruber et al., 2013), we draw on prior qualitative work to conceptualize knowledge production as a production process that involves a number of functional activities. This approach allows us to consider that team members may play different roles and that different activities have different degrees of interdependencies with each other. Moreover, this conceptualization aligns with a large body of prior work in organizational theory, allowing us to draw on concepts such as benefits from specialization, interdependencies among activities, and decomposability of activities to develop a deeper understanding of potential drivers of division of labor in teams. We believe that the "bridging" between the study of science and the organizational literature may provide a fertile ground for future research. Among others, an important next step would be to link measures of the functional division of labor among team members to the quantity or quality of scientific output. Our framework may also be useful in thinking about potential changes in the division of labor that may result from broader trends such as the increasing role of capital intensive equipment or the diffusion of information technologies (Agrawal & Goldfarb, 2008; Stephan, 2012; Franzoni & Sauermann, 2014). Finally, it would be interesting to explore potential complementarities between our activity-based perspective and the view focusing on the recombination of knowledge. For example, since the division of labor may be quite different from the partitioning of knowledge (Postrel, 2002; Takeishi, 2002), it would be interesting to explore whether and how different activities are tied to different types of knowledge. Similarly, such an integration may also help us understand which particular team members are likely to perform which (sets of) activities, an important question that is beyond the scope of the current paper.

Our study should also be of interest to the broader organizational literature. We offer a framework that systematically discusses division of labor in teams from different perspectives and suggests a number of complementary measures. Although division of labor has long played a central role in the literature, we are not aware of prior work that highlights and integrates the different perspectives considered in our model. We hope that this framework may provide a useful tool for analyzing division of labor not just in knowledge production but also in other contexts where production processes involve a number of functional activities. By bridging between organizational studies and the study of science, our paper also considers how some of the fundamental concepts originally developed to study division of labor in the production of physical goods translate into the context of knowledge production. Our empirical results suggest that existing organizational theories can indeed provide a useful lens for studying division of labor in this increasingly important context. At the same time, this context also raises important areas for future research and conceptual broadening. For example, while conventional models of division of labor focus on the trade-off between benefits from specialization and coordination costs, knowledge production may also involve important benefits from diversity, which are often considered in the literature on knowledge production but rarely in the traditional literature on division of labor (Singh & Fleming, 2010; Gruber et al., 2013). Similarly, it may be useful to complement the conventional focus on efficiency in production with greater attention to other organizational goals such as team members' learning and professional development, an issue that becomes particularly salient when we think about division of labor in the context of academic science (Hackett, 1990). Overall, we hope that our study provides interesting conceptual and empirical insights into the division of labor in teams and that it can provide a useful foundation for future work on the organization of collaborative knowledge production.

REFERENCES

Agarwal, R., & Sonka, S. 2010. Different strokes for different folks: University programs that enable diverse career choices of young scientists. In G. Libecap (Ed.), ***Spanning Boundaries and Disciplines: University Technology Commercialization in the Idea Age***, Vol. 21: 139-164: Emerald Group Publishing Limited.

Agrawal, A., & Goldfarb, A. 2008. Restructuring research: communication costs and the democratization of university innovation. ***The American Economic Review***, 98(4): 1578-1590.

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. 2014. Publishing: Credit where credit is due. ***Nature***, 508(7496): 312-313.

Arora, A., Cohen, W. M., & Walsh, J. P. 2014. The Acquisition and Commercialization of Invention in American Manufacturing: Incidence and Impact, ***NBER Working Paper #20264***.

Barley, S. R. 1996. Technicians in the workplace: ethnographic evidence for bringing work into organization studies. ***Administrative Science Quarterly***, 41(3).

Becker, G. S., & Murphy, K. M. 1992. The division of labor, coordination costs, and knowledge. ***The Quarterly Journal of Economics***, 107(4): 1137-1160.

Bercovitz, J., & Feldman, M. 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. ***Research Policy***, 40(1): 81-93.

Bohannon, J. 2013. Who's afraid of peer review? ***Science***, 342: 60-65.

Dasgupta, P., & David, P. A. 1994. Toward a new economics of science. ***Research Policy***, 23(5): 487-521.

Delamont, S., & Atkinson, P. 2001. Doctoring uncertainty: Mastering craft knowledge. ***Social Studies of Science***, 31(1): 87-107.

Fayard, A.-L., & Metiu, A. forthcoming. The role of writing in distributed collaboration. ***Organization Science***.

Franzoni, C., & Sauermann, H. 2014. Crowd Science: The organization of scientific research in open collaborative projects. ***Research Policy***, 43(1): 1-20.

Freeman, R., Weinstein, E., Marincola, E., Rosenbaum, J., & Solomon, F. 2001. Competition and careers in biosciences. ***Science***, 294(5550): 2293.

Fuhrmann, C., Halme, D., O'Sullivan, P., & Lindstaedt, B. 2011. Improving graduate education to support a branching career pipeline: Recommendations based on a survey of doctoral students in the basic biomedical sciences. ***CBE-Life Sciences Education***, 10(3): 239-249.

Gans, J., & Murray, F. 2013. Credit history: The changing nature of scientific credit, ***NBER Working Paper #19538***.

Gibbs, J. P., & Poston, D. L. 1975. The division of labor: Conceptualization and related measures. ***Social Forces***, 53(3): 468-476.

Gruber, M., Harhoff, D., & Hoisl, K. 2013. Knowledge Recombination Across Technological Boundaries: Scientists vs. Engineers. ***Management Science***, 59(4): 837-851.

Hackett, E. J. 1990. Science as a vocation in the 1990s. ***Journal of Higher Education***, 61(3): 241-279.

Haeussler, C., & Sauermann, H. 2013. Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. ***Research Policy***, 42(3): 688-703.

Hamilton, B. H., Nickerson, J. A., & Owan, H. 2003. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. ***Journal of Political Economy***, 111(3): 465-497.

ICMJE. 2010. Uniform requirements for manuscripts submitted to biomedical journals: Ethical considerations in the conduct and reporting of research: Authorship and Contributorship, Vol. 2010.

Ivaniš, A., Hren, D., Marušić, M., & Marušić, A. 2011. Less work, less respect: authors' perceived importance of research contributions and their declared contributions to research articles. ***PLoS ONE***, 6(6): e20206.

Jones, B. 2009. The burden of knowledge and the “death of the renaissance man”: is innovation getting harder? ***Review of Economic Studies***, 76(1): 283-317.
Jones, B. F., & Weinberg, B. A. 2011. Age dynamics in scientific creativity. ***Proceedings of the National Academy of Sciences***, 108(47): 18910-18914.
Kamo, Y. 2000. “He Said, She Said”: Assessing Discrepancies in Husbands' and Wives' Reports on the Division of Household Labor. ***Social Science Research***, 29(4): 459-476.
Knorr-Cetina, K. 1999. ***Epistemic Cultures: How the Sciences Make Knowledge***. Cambridge, MA: Harvard University Press.
Latour, B., & Woolgar, S. 1979. ***Laboratory Life: The Social Construction of Scientific Facts***. Beverly Hills: Sage.
Laudel, G. 2002. Collaboration and reward: What do we measure by co-authorships? ***Research Evaluation***, 11(1): 3-15.
Lissoni, F., Montobbio, F., & Zirulia, L. 2013. Inventorship and authorship as attribution rights: An enquiry into the economics of scientific credit. ***Journal of Economic Behavior & Organization***, 95: 49-69.
Marušić, A., Bošnjak, L., & Jerončić, A. 2011. A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. ***PLoS ONE***, 6(9): e23477.
Merton, R. K. 1973. ***The Sociology of Science: Theoretical and Empirical Investigations***. Chicago: University of Chicago Press.
Nielsen, M. 2011. ***Reinventing Discovery: The New Era of Networked Science***: Princeton University Press.
Osterloh, M., & Frey, B. S. 2000. Motivation, knowledge transfer, and organizational forms. ***Organization Science***, 11(5): 538-550.
Owen-Smith, J. 2001. Managing laboratory work through skepticism: Processes of evaluation and control. ***American Sociological Review***, 66(3): 427-452.
Polanyi, M. 1966. ***The Tacit Dimension***: University of Chicago Press.
Postrel, S. 2002. Islands of shared knowledge: Specialization and mutual understanding in problem-solving teams. ***Organization Science***, 13(3): 303-320.
Shibayama, S., Baba, Y., & Walsh, J. P. 2013. Adapting laboratories in biology: How techno-social contexts determine task allocation, ***Working Paper***.
Simon, H. A. 1962. The architecture of complexity. ***Proceedings of the American Philosophical Society***: 467-482.
Simonton, D. K. 2003. Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. ***Psychological Bulletin***, 129(4): 475-494.
Singh, J., & Fleming, L. 2010. Lone inventors as sources of breakthroughs: Myth or reality? ***Management Science***, 56(1): 41-56.
Smith, A. 1776. ***An Inquiry into the Nature and Causes of the Wealth of Nations***.
Stephan, P. 2012. ***How Economics Shapes Science***: Harvard University Press.
Stigler, G. J. 1951. The division of labor is limited by the extent of the market. ***The Journal of Political Economy***: 185-193.
Takeishi, A. 2002. Knowledge partitioning in the interfirm division of labor: The case of automotive product development. ***Organization Science***, 13(3): 321-338.
Teodoridis, F. 2013. Generalists, specialists, and the direction of inventive activity, ***Working Paper***.
Thompson, J. D. 1967. ***Organizations in Action***.
Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. 1976. Determinants of coordination modes within organizations. ***American Sociological Review***: 322-338.
Van Knippenberg, D., De Dreu, C. K., & Homan, A. C. 2004. Work group diversity and group performance: an integrative model and research agenda. ***Journal of Applied Psychology***, 89(6): 1008.
Van Tienderen, P. H. 1991. Evolution of generalists and specialist in spatially heterogeneous environments. ***Evolution***, 45(6): 1317-1331.

Venkatraman, V. 2010. Conventions of scientific authorship. ***Science Careers***, 4.
Von Hippel, E. 1990. Task partitioning: An innovation process variable. ***Research Policy***, 19(5): 407-418.
Wuchty, S., Jones, B., & Uzzi, B. 2007. The increasing dominance of teams in the production of knowledge. ***Science***, 316(5827): 1036-1039.

Table 2: Descriptive statistics

	Mean	SD	Min	Max
Author level (n=91311)				
conceived	0.48	0.50	0	1
performed	0.51	0.50	0	1
materials	0.34	0.47	0	1
analyzed	0.58	0.49	0	1
wrote	0.47	0.50	0	1
other contribution	0.09	0.28	0	1
ln_priorpubs_quantity	1.15	1.45	0	5.30
ln_priorpubs_quality	0.72	0.90	0	6.06
professionalage	7.65	11.04	0	60
institutionrank	5.62	2.69	1	8
developing country	0.14	0.34	0	1
missingscopus	0.52	0.50	0	1
Team level (n=13935)				
t_avg_shareofcontributions	0.56	0.17	0.18	1
t_stdev_shareofcontributions	1.05	0.44	0	2.83
t_shareconceived	0.54	0.28	0	1
t_shareperformed	0.51	0.25	0	1
t_sharematerials	0.33	0.31	0	1
t_shareanalyzed	0.62	0.28	0	1
t_sharewrote	0.55	0.30	0.05	1
t_shareother	0.09	0.22	0	1
t_teamsize	6.55	3.58	2	20
t_country_US	0.38	0.48	0	1
t_country_developing	0.11	0.31	0	1
t_firm	0.14	0.35	0	1
t_affiliations	3.34	2.08	1	23
t_funding_basic	0.31	0.46	0	1
t_funding_industry	0.05	0.22	0	1
t_funding_other	0.61	0.49	0	1
t_fieldcount	2.49	1.18	0	14
t_pages	9.96	3.29	2	51
t_publication_date	18128.74	362.51	17204	18878
t_totalactivitieslisted	4.80	0.68	1	6

Table 3: Relationships between activities

VARIABLES	(1) wrote	(2) analyzed	(3) materials	(4) performed	(5) conceived	(6) othercontribution
wrote		0.301**	-0.022**	-0.078**	0.490**	-0.043**
		(0.005)	(0.005)	(0.005)	(0.004)	(0.002)
analyzed	0.241**		-0.083**	0.117**	0.067**	-0.021**
	(0.004)		(0.004)	(0.005)	(0.004)	(0.002)
materials	-0.019**	-0.091**		-0.234**	-0.026**	-0.024**
	(0.004)	(0.005)		(0.005)	(0.005)	(0.002)
performed	-0.054**	0.103**	-0.186**		-0.048**	-0.052**
	(0.004)	(0.004)	(0.004)		(0.004)	(0.002)
conceived	0.424**	0.072**	-0.026**	-0.059**		-0.010**
	(0.004)	(0.005)	(0.004)	(0.005)		(0.002)
othercontribution	-0.157**	-0.098**	-0.101**	-0.273**	-0.041**	
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	
priorpubs_quantity	0.001	-0.008**	0.008**	-0.066**	0.027**	0.004**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
priorpubs_quality	0.011**	0.006	0.011**	-0.040**	0.024**	-0.002
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)
professionalage	0.000*	-0.004**	0.001**	-0.009**	0.002**	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
institutionrank	-0.008**	-0.005**	0.006**	0.000	-0.005**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
country_developing	-0.028**	-0.078**	0.034**	0.116**	0.018	0.013*
	(0.009)	(0.010)	(0.010)	(0.011)	(0.010)	(0.005)
missingscopus	-0.008	-0.090**	0.020**	-0.277**	0.076**	0.005
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.003)
Constant	0.213**	0.513**	0.448**	0.908**	0.170**	0.151**
	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.005)
Observations	91,311	91,311	91,311	91,311	91,311	91,311
R-squared	0.326	0.159	0.069	0.151	0.276	0.035
Number of t_paperid	13,935	13,935	13,935	13,935	13,935	13,935
r2	0.326	0.159	0.0691	0.151	0.276	0.0346

Note: OLS with paper fixed effects. *=significant at 5%, **=significant at 1%.

Table 4: Division of labor and team size

	1	2	3	4	5	6	7	8
	t_avg_ shareofcontributions	t_stdev_ shareofcontributions	t_shareconceived	t_shareperformed	t_sharecontributed	t_shareanalyzed	t_sharewrote	t_shareother
t_teamsize: 3	-0.106**	0.037**	-0.132**	-0.043**	-0.142**	-0.085**	-0.149**	-0.071*
	(0.005)	(0.003)	(0.009)	(0.008)	(0.013)	(0.010)	(0.009)	(0.030)
t_teamsize: 4	-0.166**	0.042**	-0.228**	-0.046**	-0.203**	-0.127**	-0.250**	-0.097**
	(0.005)	(0.003)	(0.009)	(0.008)	(0.013)	(0.010)	(0.009)	(0.030)
t_teamsize: 5	-0.224**	0.043**	-0.304**	-0.057**	-0.265**	-0.179**	-0.337**	-0.121**
	(0.005)	(0.003)	(0.009)	(0.009)	(0.013)	(0.010)	(0.009)	(0.030)
t_teamsize: 6	-0.251**	0.042**	-0.345**	-0.053**	-0.276**	-0.214**	-0.392**	-0.121**
	(0.005)	(0.004)	(0.009)	(0.009)	(0.013)	(0.010)	(0.010)	(0.031)
t_teamsize: 7	-0.272**	0.038**	-0.384**	-0.045**	-0.278**	-0.232**	-0.424**	-0.218**
	(0.005)	(0.004)	(0.010)	(0.009)	(0.014)	(0.011)	(0.010)	(0.032)
t_teamsize: 8	-0.298**	0.032**	-0.408**	-0.052**	-0.326**	-0.248**	-0.473**	-0.200**
	(0.006)	(0.004)	(0.011)	(0.010)	(0.014)	(0.012)	(0.011)	(0.032)
t_teamsize: 9	-0.308**	0.031**	-0.447**	-0.033**	-0.311**	-0.261**	-0.502**	-0.176**
	(0.006)	(0.004)	(0.011)	(0.011)	(0.015)	(0.012)	(0.011)	(0.033)
t_teamsize: 10	-0.317**	0.024**	-0.445**	-0.028*	-0.330**	-0.274**	-0.516**	-0.223**
	(0.007)	(0.005)	(0.012)	(0.011)	(0.016)	(0.013)	(0.012)	(0.036)
t_teamsize: 11	-0.334**	0.019**	-0.473**	-0.035**	-0.338**	-0.284**	-0.547**	-0.207**
	(0.007)	(0.005)	(0.013)	(0.013)	(0.017)	(0.015)	(0.014)	(0.038)
t_teamsize: 12	-0.341**	0.018**	-0.496**	-0.005	-0.345**	-0.282**	-0.561**	-0.328**
	(0.008)	(0.006)	(0.015)	(0.014)	(0.018)	(0.016)	(0.015)	(0.041)
t_teamsize: 13	-0.347**	0.017**	-0.513**	-0.002	-0.360**	-0.283**	-0.586**	-0.236**
	(0.009)	(0.006)	(0.017)	(0.016)	(0.021)	(0.018)	(0.017)	(0.044)
t_teamsize: 14	-0.349**	0.019**	-0.518**	0.000	-0.341**	-0.292**	-0.594**	-0.271**
	(0.010)	(0.007)	(0.019)	(0.017)	(0.023)	(0.021)	(0.019)	(0.046)
t_teamsize: 15	-0.376**	0.013	-0.518**	0.005	-0.353**	-0.381**	-0.637**	-0.259**
	(0.012)	(0.008)	(0.021)	(0.020)	(0.025)	(0.023)	(0.022)	(0.052)
t_teamsize: 16	-0.351**	0.011	-0.505**	0.037	-0.376**	-0.318**	-0.584**	-0.321**
	(0.012)	(0.009)	(0.022)	(0.021)	(0.027)	(0.025)	(0.023)	(0.060)
t_teamsize: 17	-0.363**	0.010	-0.499**	-0.001	-0.371**	-0.311**	-0.669**	-0.234**
	(0.016)	(0.011)	(0.030)	(0.028)	(0.035)	(0.033)	(0.030)	(0.064)
t_teamsize: 18	-0.358**	0.005	-0.553**	0.003	-0.355**	-0.314**	-0.580**	-0.148
	(0.017)	(0.012)	(0.031)	(0.029)	(0.036)	(0.034)	(0.032)	(0.085)
t_teamsize: 19	-0.335**	-0.002	-0.490**	0.073*	-0.338**	-0.258**	-0.646**	-0.307**
	(0.020)	(0.014)	(0.036)	(0.033)	(0.042)	(0.040)	(0.037)	(0.089)
t_teamsize: 20	-0.356**	0.009	-0.542**	0.057	-0.315**	-0.302**	-0.647**	-0.383**
	(0.021)	(0.015)	(0.039)	(0.037)	(0.045)	(0.043)	(0.040)	(0.089)
t_firm	0.001	-0.006**	-0.001	-0.001	0.005	0.017**	-0.010	-0.006
	(0.003)	(0.002)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.015)
t_funding_basic	0.010	-0.006	0.013	0.011	0.002	0.012	0.002	0.035
	(0.005)	(0.004)	(0.010)	(0.009)	(0.013)	(0.011)	(0.010)	(0.023)
t_funding_industry	0.009	0.003	0.001	0.019	-0.009	0.006	0.013	0.043
	(0.006)	(0.004)	(0.011)	(0.011)	(0.014)	(0.012)	(0.011)	(0.026)
t_funding_other	0.002	-0.004	-0.004	0.010	0.002	0.003	-0.010	0.037
	(0.005)	(0.004)	(0.010)	(0.009)	(0.013)	(0.010)	(0.010)	(0.023)
t_affiliations	0.004**	-0.000	0.008**	-0.018**	0.013**	-0.002	0.015**	0.017**
	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
t_country_US	-0.000	0.006**	-0.022**	0.021**	0.001	0.014**	-0.009	-0.037**
	(0.003)	(0.002)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.013)
t_country_developing	-0.020**	0.002	-0.043**	0.002	-0.011	-0.013	-0.026**	-0.043*
	(0.004)	(0.003)	(0.007)	(0.006)	(0.008)	(0.007)	(0.007)	(0.017)
t_totalactivitieslisted	-0.056**	-0.005**	-0.031**	-0.060**	-0.034**	-0.041**	-0.045**	-0.089**
	(0.002)	(0.001)	(0.003)	(0.003)	(0.006)	(0.004)	(0.003)	(0.007)
t_pages	0.002**	0.000	-0.001*	0.005**	0.000	0.006**	-0.002**	0.002
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
t_publication_date	0.000**	-0.000	0.000**	0.000	0.000	0.000*	0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Field	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	0.793**	0.282**	0.645**	0.838**	0.684**	0.673**	0.675**	1.340**
	(0.056)	(0.039)	(0.102)	(0.096)	(0.135)	(0.112)	(0.104)	(0.251)
Observations	13,935	13,935	13,715	13,207	9,745	13,786	13,935	2,478
R-squared	0.450	0.044	0.300	0.109	0.123	0.148	0.370	0.215

Note: OLS. *=significant at 5%, **=significant at 1%

Table 5: Differences in division of labor across contexts and types of projects

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	t_avg_ shareofcontributions	t_share conceived	t_share performed	t_share contributed	t_share analyzed	t_share wrote	t_shareother	t_avg_ shareofcontributions	t_share conceived	t_share performed	t_share contributed	t_share analyzed	t_share wrote	t_shareother
t_field_new								-0.013**	-0.011*	-0.051**	0.003	-0.008	0.005	-0.021
								(0.003)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.013)
t_interdisciplinarity_2or3fields								0.005	-0.000	0.013**	-0.011	0.038**	-0.014*	-0.013
								(0.003)	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	(0.014)
t_interdisciplinarity_3ormorefields								0.017**	0.010	0.030**	-0.001	0.058**	-0.008	-0.012
								(0.004)	(0.007)	(0.007)	(0.009)	(0.008)	(0.007)	(0.018)
biology	-0.016**	-0.033**	0.003	-0.014	0.026**	-0.050**	-0.047**							
	(0.003)	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)	(0.014)							
physical sciences	0.012	0.014	-0.039**	0.056**	-0.018	0.043**	0.029							
	(0.006)	(0.011)	(0.011)	(0.015)	(0.012)	(0.011)	(0.025)							
earth sciences	0.034**	0.017	0.041**	0.070**	-0.015	0.064**	0.106**							
	(0.009)	(0.016)	(0.015)	(0.020)	(0.017)	(0.016)	(0.041)							
computer/information sciences	0.017	0.036*	-0.010	0.087**	-0.038	0.041*	-0.007							
	(0.010)	(0.018)	(0.017)	(0.023)	(0.019)	(0.018)	(0.043)							
medicine/health	0.005*	-0.006	0.005	-0.004	0.033**	-0.006	0.012							
	(0.003)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)	(0.013)							
other	0.005	0.021	-0.034	-0.083	0.049	0.067	0.046							
	(0.025)	(0.046)	(0.044)	(0.067)	(0.051)	(0.048)	(0.130)							
t_firm	0.000	-0.001	-0.003	0.003	0.017*	-0.011	-0.010	-0.000	-0.003	-0.001	0.003	0.016*	-0.014*	-0.010
	(0.003)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.015)	(0.003)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.015)
t_funding_basic	0.008	0.011	0.018	-0.007	0.021*	-0.013	0.028	0.007	0.010	0.018	-0.009	0.018	-0.010	0.036
	(0.005)	(0.010)	(0.010)	(0.013)	(0.011)	(0.010)	(0.023)	(0.005)	(0.010)	(0.010)	(0.013)	(0.011)	(0.010)	(0.023)
t_funding_industry	0.008	0.001	0.019	-0.013	0.008	0.008	0.043	0.008	0.001	0.020	-0.015	0.007	0.011	0.052*
	(0.006)	(0.011)	(0.011)	(0.014)	(0.012)	(0.012)	(0.026)	(0.006)	(0.011)	(0.011)	(0.014)	(0.012)	(0.012)	(0.026)
t_funding_other	0.000	-0.005	0.016	-0.006	0.008	-0.020*	0.036	-0.001	-0.007	0.016	-0.008	0.006	-0.020*	0.043
	(0.005)	(0.010)	(0.009)	(0.013)	(0.011)	(0.010)	(0.023)	(0.005)	(0.010)	(0.009)	(0.013)	(0.011)	(0.010)	(0.023)
t_affiliations	0.004**	0.009**	-0.020**	0.016**	-0.006**	0.019**	0.020**	0.004**	0.009**	-0.020**	0.016**	-0.005**	0.019**	0.019**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
t_country_US	-0.001	-0.022**	0.016**	0.002	0.015**	-0.008	-0.034**	-0.001	-0.022**	0.018**	-0.001	0.017**	-0.011*	-0.039**
	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.013)	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.013)
t_country_developing	-0.018**	-0.039**	-0.005	-0.008	-0.012	-0.015*	-0.036*	-0.019**	-0.039**	-0.002	-0.012	-0.010	-0.023**	-0.045**
	(0.004)	(0.007)	(0.006)	(0.008)	(0.007)	(0.007)	(0.017)	(0.004)	(0.007)	(0.006)	(0.008)	(0.007)	(0.007)	(0.017)
t_totalactivitieslisted	-0.056**	-0.032**	-0.061**	-0.032**	-0.038**	-0.048**	-0.096**	-0.056**	-0.033**	-0.062**	-0.032**	-0.039**	-0.047**	-0.093**
	(0.002)	(0.003)	(0.003)	(0.006)	(0.004)	(0.003)	(0.006)	(0.002)	(0.003)	(0.003)	(0.006)	(0.004)	(0.003)	(0.006)
t_pages	0.001**	-0.002**	0.005**	-0.001	0.008**	-0.004**	-0.000	0.001**	-0.002**	0.006**	-0.001	0.007**	-0.003**	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
t_publication_date	0.000**	0.000**	0.000	0.000	0.000*	0.000**	-0.000	0.000**	0.000**	0.000	0.000	0.000*	0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team size dummies	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	0.790**	0.669**	0.802**	0.730**	0.589**	0.747**	1.361**	0.926**	0.786**	0.795**	0.930**	0.424**	1.209**	1.811**
	(0.056)	(0.101)	(0.097)	(0.134)	(0.113)	(0.105)	(0.249)	(0.060)	(0.109)	(0.105)	(0.144)	(0.121)	(0.112)	(0.265)
Observations	13,935	13,715	13,207	9,745	13,786	13,935	2,478	13,909	13,689	13,182	9,731	13,760	13,909	2,474
R-squared	0.442	0.295	0.080	0.113	0.120	0.352	0.190	0.444	0.294	0.086	0.111	0.121	0.356	0.194

Note: OLS. *=significant at 5%, **=significant at 1%