Science Policy Research Report: Funding Team Science

Jonathon N. Cummings
Duke University

August 31, 2018

An earlier version of this report was presented at the January 8-9, 2018
Workshop on Government Decision-Making to Allocate Scientific Resources
(National Academies Keck Center)

Please contact me (jonathon.cummings@duke.edu) with any missing references on
science team process and/or performance as they relate to science team size

The writing of this report was supported by an award from the
National Science Foundation (SciSIP #1735701)

Science Policy Research Report: Funding Team Science
Executive Summary

Team science is the collaboration of two or more scientists working interdependently towards a common research goal. As funding agencies, such as the National Science Foundation and the National Institutes of Health, continue to promote team science, it is important to evaluate how team size impacts team process (how the team works together) and team performance (what the team produces). Given scarce financial resources, an important science policy issue is how to effectively allocate funding to teams of varying sizes. For example, given a remaining program budget of US$1.5M, when is it better to fund one larger team of 15 scientists ($1.5M for entire team) versus three smaller teams of five scientists ($500,000 for each team)? And should the funding decision differ for projects with low versus high complexity (e.g., interdisciplinary, multi-institutional, and/or variety of interdependent sub-tasks required by the work)?

Through a review of published empirical research on science teams, this report explores the benefits and costs of funding teams of varying sizes. In addition to synthesizing the literature on team size in science, the analysis also includes recent research on science team complexity. A distinction is made in the literature between smaller-scale studies (e.g., dozens to hundreds of teams) and larger-scale studies (e.g., thousands to millions of teams). Furthermore, to assess whether the evidence identified is robust and reliable, three categories are used to classify prior empirical work on science teams: (a) strong causal inference (e.g., laboratory experiments and field experiments), (b) moderate causal inference (e.g., longitudinal studies with appropriate controls), and (c) weak causal inference (e.g., cross-sectional surveys and observational studies).

Empirical findings from the literature review on science team size highlight a puzzle: While *smaller-scale studies* on science teams primarily demonstrate a negative relationship between team size and process (e.g., larger teams, on average, report being less well-coordinated than smaller teams), *larger-scale studies* on science teams primarily demonstrate a positive relationship between team size and performance (e.g., larger teams produce more publications, patents, and citations than smaller teams). The implication is that while funding larger teams can make research integration more difficult (especially teams high in complexity), it can also increase member effort, expertise, and division of labor. However, since larger-scale studies on science teams have focused on performance (and not on process), future research should focus on *why* team size is positively related to performance in science teams.

Based on the evidence, policy recommendations are provided to guide the allocation of funding to team science in order to facilitate complex problem solving, new research discoveries, and technological innovation. Specific policy recommendations for funding agencies include (1) adding team size as a factor in the evaluation of proposals, (2) adding research integration as a factor in the evaluation of proposals with high complexity, and (3) running experiments to assess how evaluating team size and research integration impacts team performance. Suggestions for implementing these recommendations include: (1a) requiring PIs to justify team size in proposals, (1b) requiring reviewers to evaluate team size justification, (2a) requiring PIs to include research integration plans when team complexity is high, (2b) requiring reviewers to evaluate research integration plans when team complexity is high, and (3) randomly assigning proposals to panels in which team size and research integration are evaluated or not evaluated.

Science Policy Research Report: Funding Team Science

"…while the increasing size of team-based research projects brings greater scientific expertise and more advanced instrumentation to a research question, it also increases the time required for communication and coordination of work (pg. 1)."

-- *Enhancing the Effectiveness of Team Science* (National Academies Press, 2015)

Imagine you are a funding agency program officer with US$1.5M remaining in your budget. You have 4 proposals on your desk, all with the highest possible ratings. One proposal has a 15-member team. Each of the other three proposals has a 5-member team. Furthermore, the proposed science for the one 15-member team is identical to the proposed science across the three 5-member teams. All else equal, when should you spend your remaining budget on the 15-member team, and when should you spend it on the three 5-member teams? Should your decision depend on team complexity – number of member disciplines, member institutions, or variety of interdependent sub-tasks required by the work? This research report explores the science policy implications of funding teams of scientists. In particular, it focuses on what the empirical evidence says about the tradeoffs of funding larger versus smaller science teams.

Teams are typically defined as the collaboration of two or more members working interdependently towards a common goal (Borner, et al. 2010; Fiore, 2008; Guimera, et al., 2005; Hackman, 1987; Katz & Martin, 1997; Kozlowski & Ilgen, 2006; Stokols, et al., 2008). The science itself conducted by teams can range from basic to applied science, from engineering to technology development, or from invention to product extension. Team science can take place in universities, corporate labs, government organizations, and other institutions engaged in science (National Academies Press, 2015).

A steady increase in the size of science teams over the past 50 years has resulted in renewed interest in the topic of team size (Hudson, 1996; Lariviere, et al., 2015; Leahey, 2016; Petersen, Pavlidis, & Semendeferi, 2014). For example, Wuchty, Jones, and Uzzi (2007) analyzed close to 20 million publications and reported (a) an increase in the percentage of papers with two or more authors in Science & Engineering (from 50% in 1960 to 80% in 2000) and (b) an increase in the number of publication authors in Science & Engineering (going from a mean of 2 authors in 1960 to a mean of 3.5 authors in 2000). By 2013, the percentage of Science & Engineering publications with 2 or more authors had reached 90%. While some disciplines, such as physics (e.g., Large Hadron Collider; Merali, 2010), astronomy (Planck observatory; Seljak, 2012), and genomics (The ENCODE Project Consortium, 2012), have successfully deployed large "teams" of 100s or 1000s of members, it is important to note that 95% of Science & Engineering publications still have 10 or fewer authors (National Academies Press, 2015).

Public and private investment have played a critical role in the advancement of team science (Stokols et al., 2008). Funding agency program managers want to allocate resources to optimize complex problem solving, new research discoveries, and technological innovation. Resource allocation includes determining which teams should get funded and how much funding they should get. Traditional funding mechanisms (e.g., NSF Core Program awards or NIH Research Project Grant Program [R01] awards) as well as special programs (e.g., DARPA Grand

Challenges or NSF/NIH Brain Initiative) have encouraged larger teams across academic institutions to tackle broader, more complex, larger-scale, and/or inter-disciplinary research questions (Croyle, 2012; Cummings & Kiesler, 2005; Dahlander & McFarland, 2013; Defazio, Lockett, & Wright, 2009; Jones, Wuchty, & Uzzi, 2008; Wadman, 2010).

Scope of Teams Considered in the Review of Scientific Evidence

Given that a range of scientific disciplines have studied the topic of team science (e.g., organizational behavior, psychology, sociology, economics, science studies, research policy, computer science, physics, social science & medicine), an inter-disciplinary review of the empirical findings is required to create a deeper understanding of the impact of team size in team science. In particular, Google Scholar (with search terms such as "team science," "team size," "group size," "number of members," "number of authors," and "number of inventors") was used to identify relevant published articles on team science size (see References). In addition, for the identified articles, forward and backward citations of these articles were examined.

As the typical size of teams studied in the academic literature ranges from 2 to roughly 15 members (National Academies Press, 2015), this review does not consider multi-team systems (e.g., multiple teams of experts; see Asencio, Carter, DeChurch, Zaccaro, & Fiore, 2012), research centers (e.g., NSF Engineering Research Centers; see Boardman & Corley, 2008; NIH transdisciplinary research center grants; see Hall et al., 2012), scientific networks (international collaborations; see Wagner & Leydesdorff, 2005), online communities (e.g., Wikipedia; see Zhang & Zhu, 2011), crowdsourcing (e.g., MTurk; see Mao, Mason, Suri, & Watts, 2016), social dilemmas (e.g., collective goods; see Chamberlin, 1974) or other groups larger than 15 members. Moreover, this review does not address prior literature on individual productivity (e.g., Foster, Rzhetsky, & Evans, 2015; Leahey, Beckman, & Stanko, 2017; Lee & Bozeman, 2005) or institutional productivity (e.g., Shibayama, Baba, & Walsh, 2015; Stahl, Leap, & Wei, 1988).

Since the focus is on science teams, it is important to note how these teams compare to non-science teams that work outside of a scientific context. For example, science teams are more similar to non-science teams that perform intellectual tasks (McGrath, 1984) than to non-science teams that perform physical tasks such as pulling on a rope (e.g., Ringelmann effect; Ingham, Levinger, Graves, & Peckham, 1974) or clapping and shouting (e.g., social loafing; Latane, Williams, & Harkins, 1979). Science teams are also more similar in terms of scope of work and time frame to 'real teams' (e.g., software development teams, corporate product teams, health care teams) than 'nominal groups' (e.g., brainstorming groups in the lab; see Diehl & Stroebe, 1987). Finally, science teams operate as interacting groups in which members have influence over one another through interaction (e.g., behaviors in the presence of one other, creating products for one another, or communicating with one another; see Kelley & Thibaut, 1954) rather than non-interacting groups in which members combine opinions or judgments but do not actually interact (e.g., wisdom of crowds; see Larrick & Soll, 2006).

Evaluating robust and reliable evidence

In order to synthesize the academic literature on team size in science teams it is important to evaluate the evidence before providing recommendations for policy and practice. In general, the stronger the evidence, the more specific the recommendations can be (and the weaker the evidence, the broader the recommendations should be). Three categories are used to classify the body of evidence on team size and team complexity in science teams:

(1) Strong causal inference (e.g., laboratory experiments and field experiments)

(2) Moderate causal inference (e.g., longitudinal studies with appropriate controls, such as prior individual experience and prior team experience)

(3) Weak causal inference (e.g., cross-sectional surveys and observational studies)

It is worth noting that no previously published studies on team science could be classified as providing strong causal inference about the impact of team size on team process or team performance. For example, there are not any randomized controlled trials (RCTs) on how varying science team size impacts how well scientists work together or how well they perform. And even the larger-scale longitudinal studies providing moderate causal inference still suffer from an important methodological limitation around sample selection bias. This notable challenge to robust and reliable evidence is discussed in more detail below.

Sample Selection Bias. A popular approach to studying team science is through an analysis of publication authors, where all of the authors on a publication represent a "science team" (e.g., Adams, Black, Clemmons, & Stephan, 2005; Guimera, Uzzi, Spiro, & Amaral, 2005; Lee, Walsh, & Wang, 2015; Milojevic, 2014; Wu, Wang, & Evans, 2018). The set of publications becomes a sample, therefore any team that did not publish is not part of the sample. As a result, the sample of publications is composed of successful science teams, and does not represent the population of science teams more generally. To illustrate the selection bias – e.g., how a sample of science teams that were successful in publishing might be different from a sample of science teams that were unsuccessful in publishing – consider hypothetical stages of development in science teams.

Receive request for proposal (RFP): Some teams form right away to start working on a grant proposal request and other teams form but do not start working on a grant proposal request. Are successful teams more likely to start working on a grant proposal request than unsuccessful teams? If yes, then successful teams would differ from unsuccessful teams in ways related to team formation (e.g., better fit between member interests and the RFP, and potentially better team process working together such as getting off to a strong start).

Apply for a grant: Some teams that formed end up submitting a proposal and other teams do not. Are successful teams more likely to actually apply for a grant than unsuccessful teams? If yes, then successful teams would differ from unsuccessful teams in ways related to completing the proposal (e.g., better fit between member roles and completing the proposal, and potentially better team process working together such as efficiently dividing up who does what).

Receiving the grant funding: Some teams that submitted a grant proposal receive the funding and other teams do not. Are successful teams more likely to receive grant funding than unsuccessful teams? If yes, then successful teams would differ from unsuccessful teams in ways related to proposal quality (e.g., better fit between member expertise and proposal quality, and potentially better team process working together such as generating and refining creative ideas).

Conducting research together: Some teams that receive grant funding conduct their research together and other teams split up the funding to conduct their research separately. Are successful teams more likely to conduct research together than unsuccessful teams? If yes, then successful teams would differ from unsuccessful teams in ways related to research integration (e.g., better fit between member complementarities and research integration, and potentially better team process working together such as smoothly communicating across different knowledge areas).

Writing papers: Some teams conducting research together write papers based on the funded proposal and other teams do not (e.g., if there are null results from the funded research). Are successful teams more likely to write papers based on a funded proposal than unsuccessful teams? If yes, then successful teams would differ from unsuccessful teams in ways related to research output (e.g., better fit between member execution and research output, and potentially better team process working together such as effectively writing up studies together).

Publishing a paper: Some teams that write papers get at least one paper published and other teams do not get any papers published. By definition, successful teams publish at least one paper in order to appear in the sample. If the odds of publishing increase when forming a team after receiving a request for proposal, actually applying for a grant, receiving the grant funding, conducting research together, and writing papers based on the funded proposal, then at each stage of development, successful teams are likely to differ from unsuccessful teams.

The sample selection bias is not limited to selecting the sample based on a dependent variable (i.e., science teams that were successful in publishing at least one paper). There are also concerns about the measure of team size in the sample of successful teams, and how the actual number of members in the sample of successful teams differed from science teams more generally. For example, what if a 5-member team formed in response to the RFP dropped two existing members before applying for the grant, added two new members as a condition for funding, dropped two members when conducting the research, and added two new members when writing up a paper that was published. A 5-member team (in the case of 5 authors on a publication) could have had 9 members, which is a notably larger team.

The implication of team size at the end of a project being different than team size at the beginning of a project extends further if a sample of successful teams differs from a sample of unsuccessful team in terms of *why* members joined or left a team. Members could have been added or dropped because of endogenous reasons (e.g., a team member wanted to bring in a friend from another project, or a team member wanted to leave because the project did not seem promising) or exogenous reasons (e.g., the program officer wanted the team to add an expert on a particular topic, or the proposal was funded at 75% of the requested budget and a member was dropped). Because all of the reasons why teams grew or shrank are unobserved in the sample of successful teams who published a paper, it is possible that teams achieving more at each stage became larger (and teams achieving less at each stage became smaller). Thus the sample of successful teams could be biased in terms of how teams reached their ultimate size.

Finally, information about the publications used to measure both team size and team performance is concentrated in a few data sources, namely Google Scholar, Web of Science (Thomson Reuters/Clarivate Analytics), Scopus (Elsevier), and PubMed (National Center for Biotechnology Information). Though they vary in terms of public access (Google Scholar, PubMed) versus private access (Web of Science, Scopus), as well as the extent of coverage (e.g., peer-reviewed publications, conference proceedings, book chapters), many peer-reviewed publications will show up in all four data sources. As a result, the sample of science teams found in these data sources will differ from the population of science teams in a number of critical ways including an overrepresentation of teams that publish in English, an overrepresentation of teams from disciplines with indexed peer-reviewed journals, and an overrepresentation of teams from institutions that are in the business of producing research in peer-reviewed journals.

Team Size

"The group [should be] just large enough to do the work. If a task requires four sets of hands, then there should be four people in the group -- but no more than that (pg. 327)."

-- *The design of work teams* (Hackman, 1987)

This section synthesizes the published scientific evidence in a way that summarizes and clarifies what is known and what is not known about team size. It is important to note that team size for this synthesis is entirely based on the number of members (often authors of publications) and not the amount of resources devoted to the team (such as grant funding, equipment costs, or other expenses that might also reflect team size). In addition, performance is typically measured using outputs that are easy to count and comparable across teams like publications, patents, and citations, rather than other indicators of performance such as educational outcomes, translation into practice, and broader impacts that are more subjective (e.g., Cummings & Kiesler, 2007).

Based on the criteria described above, the review included empirical articles on science teams in which team size and team process (i.e., how the team works together) and/or team size and team performance (i.e., what the team produces) were measured. Steiner (1972) and others have noted the tradeoffs of team size in terms of process losses and performance gains. In larger teams there are potentially more *process losses* around motivation (e.g., free-riding), communication (e.g., keeping up with all members), and coordination (e.g., integrating different tasks). But in larger teams there are also potentially more *performance gains* via effort (e.g., more people working on task), expertise (e.g., solve problems more quickly), and division of labor (e.g., efficiently divide up tasks).

Findings from the literature review on team size suggest that while size and performance are positively related in science teams, size and process are negatively related in science teams. Note that papers are marked with [M] for moderate causal inference and [W] for weak causal inference to help evaluate the evidence. As a reminder, no papers were classified as having strong causal inference. In terms of team size and team performance, larger science teams produced more research papers (Cummings et al., 2013 [M]), journal publications (Adams et al., 2005 [M]; Cohen, 1991 [W]; Wuchty, Jones, & Uzzi, 2007 [M]), publication impact (e.g., Lee, Walsh, & Wang, 2015 [M]; Lariviere et al., 2015 [W]), patents granted (Bercovitz & Feldman, 2011 [M]), and patent impact (Breitzman & Thomas, 2015 [W]; Singh & Fleming, 2010 [M]) than smaller science teams. Larger teams were also more likely to make atypical combinations in terms of novelty and convention, which was associated with publication impact (Uzzi et al, 2013 [M]). Lee, Walsh, & Wang (2015 [M]) also found a (curvilinear) relation between team size and novelty. No studies indicated an overall neutral or negative relationship between team size and performance in science teams.

In terms of team size and team process, scholars have noted the bureaucratic challenges that accompany larger scientific collaborations (Chompalov, Genuth, & Shrum, 2002). For example, larger science teams reported more coordination costs, as indicated by fewer team meetings and knowledge transfer activities, than smaller science teams (Cummings & Kiesler, 2005 [W], 2007 [W]). Larger science teams also experienced more problems of misunderstanding and cultural differences (Walsh & Maloney, 2007 [W]) than smaller science teams. Given how few studies there were on science team process, it is also worth noting that larger *non-science teams* reported less satisfaction (Hackman & Vidmar, 1970), lower quality of experience (Aube, Rousseau, & Tremblay, 2011), less cohesion (Hoegl & Proserpio, 2004), lower perceived support (Mueller, 2012), lower perceived integration (Lichtenstein et al., 1997), lower rate of participation (Bray, Kerr, & Atkin, 1978), and more impersonal modes of coordination (Van de Ven, Delbecq, & Koenig, 1976) than smaller *non-science teams*.

Given the empirical findings, while funding larger teams is likely to increase member effort, expertise, and division of labor in the team, funding larger teams is also likely to make research integration more difficult. To further explore why team process is less likely to be studied than team performance, smaller-scale studies (e.g., dozens to hundreds of teams) were distinguished from larger-scale studies (e.g., thousands to millions of teams). Note that the articles are marked with [N=x] where x indicates the number of science teams used in the statistical analyses. Published larger-scale studies without measures of team process (in order of size) include Lariviere et al. (2015 [N = 32.5 Million]), Wuchty, Jones, and Uzzi (2007 [N = 20 Million]), Adams et al. (2005 [N = 2.4 Million]), Breitzman and Thomas (2015 [N = 1.8 Million]), Singh and Fleming (2010 [N = 255,378]), Lee, Walsh, and Wang (2015 [N = 1493]), and Bercovitz and Feldman (2011 [N = 1425]). In contrast, published smaller-scale studies with measures of team process (in order of size) include Cummings and Kiesler (2007 [N = 491]), Walsh and Maloney (2007 [N = 159], and Cummings and Kiesler (2005 [N = 62]). It seems clear that team process is more challenging to study in larger-scale studies, and thus team process is typically ignored.

Team Complexity

"The variable of group size should be included in theories of group behavior, distinguishing where possible the effects that result from the interaction of group size with other independent variables and the effects arising from intervening variables that are dependably and nondependably associated with size (pg. 383)."

-- *Effects of group size* (Thomas & Fink, 1963)

When science teams add members from different disciplines, members from different institutions, and members who are responsible for different sub-tasks, they grow *both* in size and in complexity. Team complexity captures the need to coordinate the work and manage dependencies among members (Malone & Crowston, 1994). For example, interdisciplinary teams have disciplinary differences in language and norms about the research process, multi-institutional teams have geographic dispersion and cultural differences across institutions, and teams with a variety of interdependent sub-tasks have differences involving various resources, tools, applications, databases, experiments, and other elements of the science. As a result of these kinds of complexity, research integration – combining the distinct expertise of members and their work into a unified whole – is more difficult (Balakrishnan et al., 2011; Salazar, Lant, Fiore, & Salas, 2012).

Compared to science teams working in one discipline at one institution on one task, science teams working interdependently across disciplines, institutions, and sub-tasks face more challenges managing coordination. As Steiner (1972) notes, there are more linkages to manage as teams with interdependent members grow exponentially larger (e.g., 2 member team, 1 linkage; 3 member team, 3 linkages; 4 member team, 6 linkages; 5 member team, 10 linkages; 6 member team, 15 linkages; 7 member team, 21 linkages; 8 member team, 28 linkages). Thus like with team size, team complexity can contribute to process losses (e.g., coordination costs/break into subgroups without integration) and performance gains (e.g., more resources, ideas, and perspectives). Size determined by the scope of the work (e.g., having different kinds of members because each member is needed to tackle the complex problem) is not the same as size determined by desired participation (e.g., having different kinds of members for the sake of having different kinds of members).

While the academic literature with published empirical articles on team science is becoming more established, the academic literature with published empirical articles on team complexity is still in its infancy. Three studies highlight research on the topic of team complexity. First, Jones, Wuchty, & Uzzi, 2008 [M]) show that multi-institutional teams are more highly cited when at least one member is from an elite institution. Second, Cummings et al. (2013 [M]) find that the marginal productivity of larger teams declined as their diversity increased in terms of (a) disciplines (as represented by the disciplines of the PIs) and (b) institutions (as represented by the universities of the PIs). Third, Lee, Walsh, & Wang (2015 [M]) examine field variety (measured by the areas of expertise represented on the team) and report a positive relationship with novelty (measured as the rareness of references cited together) but no relationship with being highly cited (measured as being in the top 1% of publications in the field that year).

Unfortunately, as with most of the studies linking team size to performance, studies linking team complexity to performance have not examined underlying team processes that could explain why more complex science teams are sometimes better (and sometimes worse) than less complex science teams. For example, how are disciplinary differences in language and norms about the research process resolved? How are geographic dispersion and cultural differences across institutions resolved? How are task differences involving various resources, tools, applications, databases, experiments, and other elements of the science resolved?

Policy Recommendations

This report reviewed the body of evidence on team size and team complexity in order to make recommendations to policy makers regarding the funding of team science. The target audience for these recommendations is policy makers who fund team science (NSF, NIH, DOD, Foundations, etc). A key assumption is that these policy makers want to allocate resources to teams in an optimal way to facilitate scientific breakthroughs and new discoveries. Resource allocation includes determining which teams should get funded and how much funding they should get.

Funding agency programs often set a budget range and time frame for proposals (e.g., some NSF Core Programs: Small Projects - up to $500,000 total budget with durations up to three years; Medium Projects - $500,001 to $1,200,000 total budget with durations up to four years; and Large Projects - $1,200,001 to $3,000,000 total budget with durations up to five years). Therefore, the policy recommendations are intended to help optimize the use of resources for science teams of varying sizes. While funding programs for centers (e.g., NSF Engineering Research Centers or Science and Technology Centers) have had extensive evaluation guidelines in the past to assess the potential return on investment, requiring coordination plans and other forms of prospective assessment for small and large teams has been limited.

Given that the evidence on previous empirical work on team size and team complexity was classified as providing weak to moderate causal inference, the specificity of recommendations is necessarily broad. Policy recommendations to consider for guiding the allocation of funding to team science include:

(1) adding team size as a factor in the evaluation of proposals (e.g., to judge potential performance gains from each member on science team)

(2) adding research integration as a factor in the evaluation of proposals with high complexity (e.g., to judge potential process losses from each additional discipline, institution, or sub-task on science team performance)

(3) running experiments to assess how evaluating team size and research integration impacts team performance (e.g., to test whether having panels assess team size and/or research integration increases the likelihood of teams performing better once they are funded)

Suggestions for implementing these recommendations include:

(1a) requiring PIs to justify team size in proposals (e.g., how each member will make an important contribution to the overall performance goals of the science team)

(1b) requiring reviewers to evaluate team size justification in proposals (e.g., how important the contribution of each member is to the overall performance goals of the science team)

(2a) requiring PIs to include research integration plans in proposals when team complexity is high (e.g., how members propose bridging disciplines, institutions, or sub-tasks)

(2b) requiring reviewers to evaluate research integration plans in proposals when team complexity is high (e.g., how feasible the proposed plans are for bridging disciplines, institutions, or sub-tasks)

(3) randomly assigning proposals to panels in which team size and research integration are evaluated or not evaluated (e.g., when a program has multiple panels, create a treatment group by having one or more panels include the evaluation of team size and/or research integration, and create a control group by having one or more panels *not* include the evaluation of team size and/or research integration).

Future Directions

"Economic and technological changes in research has stimulated the development of new forms of teamwork involving greater dependence of scientists on external authorities, greater centralization of authority in research organizations, and a complex division of labor involving professional technicians and professionals from various disciplines (pg. 241)."

-- *Traditional and modern forms of scientific teamwork* (Hagstrom, 1964)

Role of Team Process. Since prior published empirical work on science teams has focused primarily on performance (and not on process), it remains unclear *why* team size is positively related to performance. Are science teams different from other teams in how they experience process losses (e.g., fewer problems)? Are science teams different from other teams in how they benefit from performance gains (e.g., economies of scale)? Future research on science teams should examine team process in more detail, especially as it pertains to the different stages of development (e.g., applying for a grant, conducting research together, and writing papers). Teasing a part process losses (e.g., motivation vs communication vs coordination) and performance gains (e.g., effort vs expertise vs division of labor) in larger science teams is a natural place to start. Because science team process losses do not appear to undermine performance gains in larger-scale studies with archival data analyses, it suggests that incorporating the methods of smaller-scale studies (e.g., interviews, surveys) will be required to gain insight into why team size matters. Field experiments to study aspects of team science are particularly promising (Azoulay, 2012; Guinan, Boudreau, & Lakhani, 2013; Riedl & Woolley, 2017). These additional methods will likely be required to better understand the insidious effects of sample selection bias that continue to undermine a potentially robust and reliable evidence base on science team size.

Optimal Team Size. The myth of a universal optimal team size is predicated on the notion that all teams are doing the exact same task. As soon as the nature of work is considered, it becomes clear why an optimal team size for most physics, astronomy, or genomics teams is not going to be the same as an optimal team size for most psychology, sociology, or economics teams. And as the level of complexity in teams continues to increase, determining an optimal team size will depend on understanding how much interdependence is required among members. For example, an optimal size for teams with a clear division of labor strategy and low integration needed among members is going to be different than an optimal size for teams with no clear division of labor strategy and high integration needed among members. Future research will benefit from exploring the conditions under which certain kinds of teams appear to find an optimal team size (e.g., process losses are minimized and performance gains are maximized). There is also much more to learn about when an optimal team size is good for productivity, such as producing more publications, versus producing novel or atypical ideas, such as publications that differ from what has come before (Funk & Owen-Smith, 2016; Lee, Walsh, & Wang, 2015; Uzzi, et al., 2013; Wu, Wang, & Evans, 2018). Another unanswered question is who is in the best position to judge whether a team needs more or fewer members, and whether they should be added earlier or later in the development of the science team. Evidence in other team contexts suggest that members are not good at estimating the consequences of increasing the size of the team (Brooks, 1995; Pendharkar & Rodger, 2009; Staats, Milkman, & Fox, 2012). And linking back to the challenge of sample selection bias, is optimal team size impacted by endogenous versus exogenous reasons behind adding researchers to (or dropping researchers from) larger versus smaller science teams?

Member Participation. There is relatively little discussion in published empirical articles on team science about how variation in member participation could impact team process and team performance. For larger-scale archival analyses this is likely because there is not consistent information on who did what and how much time each participant spent on the team, though some journals have recently adopted the practice of indicating who did what (which may or may not be the same as time spent). More insight into why certain members participated in certain parts of the research (e.g., during grant writing, during data collection, during data analysis, during publication writing) will also inform our understanding of team process and team performance. An interesting question for future research is to what extent multiple-team membership and having more "full-time" versus "part-time" members on a team impacts process and performance (e.g., Carayol & Matt, 2004; Cummings & Haas, 2012; Hoegl & Weiss, 2016; O'Leary, Mortensen, & Woolley, 2011). On the one hand, working on a greater number of other teams at the same time could give members access to new ideas and practices in their other teams, but on the other hand, working on a greater number of other teams at the same time could create bottlenecks and coordination problems. Finally, what is good for the team might not necessarily be good for the member, especially if individual productivity does not improve as a result of teamwork (cf. Bikard, Murray, & Gans, 2015; Fanelli & Larivière, 2016; Lee & Bozeman, 2005; Seglen & Aksnes, 2000; Stallings et al., 2013). Understanding how members make tradeoffs between participating in individual and team work (and between work on large and small teams) will also shed light on the extent of sample selection bias in research on science teams.

Size as Signal. Because the performance of science teams depends on human perceptions of quality (e.g., novelty, interest, credibility) rather than entirely objective measures of quality (e.g., time, cost, productivity), we do not know if the relationship between size and performance is entirely explained by what the teams actually produce. That is, a larger team may signal higher quality because size could convey positive information about the team (e.g., ability or competence) to an audience, much like advertising a product (cf. Nelson, 1974; Spence, 1973). In the market for academic ideas, an audience of grant panelists, journal editors, and article readers may perceive higher quality when there are more science team members (e.g., "…if there are more authors, the work must be better"). Similarly, team complexity (more disciplines, more universities, and/or more interdependent sub-tasks, which is associated with higher team size) could also signal higher quality, especially if at least one of the institutions is well-known (Jones, Wuchty, & Uzzi, 2008). Thus it is important for future research to evaluate the role of signaling in larger, more complex teams. In some initial evidence from a grant proposal review process, Cummings & Kiesler (2008, pg. 1630) found that, conditional on proposal quality (as rated by reviewers in advance of panel discussion), funded proposals (based on panel ratings) were significantly larger and had significantly more universities than unfunded proposals. This suggests that sample selection bias is likely involved in putting larger, multi-institutional teams on the path for success well before the research is conducted and written up for publication.

Conclusion

As team science continues to grow in popularity for both scientists and funding agencies, it is important to take a critical view towards understanding the nature of the significant positive relationship between size and performance. Whether it is incorporating process measures into empirical studies of team science or teasing a part potential biases in favor of larger teams from the actual quality of their output, research can advance our knowledge of team science as well as shape the science policy that accompanies what is learned.

The policy recommendations based on the synthesis of published empirical results should inform government agencies (e.g., National Science Foundation, National Institutes of Health, and Department of Defense) and other organizations that fund scientific research. The recommended policies have the potential to guide the efficient allocation of resources in order to maximize the return on scientific investment. Societal benefits include funding team science in a way that facilitates complex problem solving, new research discoveries, and technological innovation.

This report began with the scenario of a funding agency program officer with US$1.5M remaining in the budget, and the question of when it is better to fund one larger team of 15 researchers ($1.5M for team) versus three smaller teams of five researchers each ($500,000 for each team). Thinking about the question from the perspective of team complexity, and the need for members to integrate across disciplines, institutions, or sub-tasks, might give different answers. If the goal is to publish new knowledge, and team complexity is low, then a larger team of 15 researchers (working relatively independently) will likely produce more publications than the three smaller teams of five researchers given that the larger team can scale up production through division of labor. If team complexity is high, then the three smaller teams of five researchers (each working relatively interdependently) will likely produce more publications than the larger team of 15 researchers given that the smaller teams incur much lower coordination costs when integrating their parts of the task.

Selected Publications

1. National Research Council. (2015). *Enhancing the effectiveness of team science*. In Cooke, N. Hilton, M., et al. (pp. 1-245). The National Academies Press: Washington, DC.

2. Stokols, D., Hall, K., Taylor, B., & Moser, R. (2008). The science of team science: Overview of the field and introduction of the supplement. *American Journal of Preventive Medicine, 35*(Suppl. 2), S77-S89.

3. Jones, B., Wuchty, S., & Uzzi, B. (2008). Multi-university research groups: Shifting impact, geography, and stratification in science. *Science, 322*(5905), 1259-1262.

4. Cummings, J. N., Kiesler, S., Zadeh, R., & Balakrishnan, A. (2013). Group heterogeneity increases the risks of large group size: A longitudinal study of productivity in research groups. *Psychological Science, 24*(6), 880-890.

5. Lee, Y-N, Walsh, J., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy, 44*(3), 684-697.

6. Adams, J., Black, G., Clemmons, J., & Stephan, P. (2005). Scientific teams and institutional collaborations: Evidence from US Universities, 1981-1999. *Research Policy, 34*(3), 259-285.

7. Leahey, E., Beckman, C., & Stanko, T. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly, 62*(1), 105-139.

8. Dahlander, L., & McFarland, D. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly, 58*(1), 69-110.

9. Bikard, M., Murray, F., & Gans, J. (2015). Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Management Science, 61*(7), 1473-95.

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

References

** included in the synthesis of published empirical results on science team size and complexity*

*Adams, J., Black, G., Clemmons, J., & Stephan, P., (2005). Scientific teams and institutional collaborations: Evidence from US Universities, 1981-1999. *Research Policy, 34*(3), 259-285.

Asencio, R., Carter, D.R., DeChurch, L.A., Zaccaro, S.J., & Fiore, S.M. (2012). Charting a course for collaboration: A multiteam perspective. *Translational and Behavioral Medicine, 2*(4), 487-494.

Aube, C., Rousseau, V., & Tremblay, S. (2011). Team size and quality of group experience: The more the merrier? *Group Dynamics, 15*(4), 357-375.

Azoulay, P. (2012). Turn the scientific method on ourselves. *Nature, 484*, 31-32.

Balakrishnan, A. D., Kiesler, S., Cummings, J. N., & Zadeh, R. (2011). Research team integration: What it is and why it matters. *Proceedings of the ACM conference on Computer-Supported Cooperative Work*, (March 19-23), Hangzhou, China.

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

Bikard, M., Murray, F., & Gans, J. (2015). Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Management Science, 61*(7), 1473-95.

Boardman, C., & E. Corley. (2008). University research centers and the composition of research collaborations. *Research Policy, 37,* 900-913.

Borner, K., et al. (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine, 2*(49), 49cm24.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology, 36*(11), 1224-1240.

*Breitzman, A., & Thomas, P. (2015). Inventor team size as a predictor of the future citation impact of patents. *Scientometrics, 103*(2), 631-647.

Brooks, F. (1995). The mythical man-month. Reading, MA: Addison-Wesley.

Carayol, N., & Matt, M. (2004). Does research organization influence academic production?: Laboratory level evidence from a large European university. *Research Policy, 33*(8), 1081-1102.

Chamberlin, J. (1974). Provision of collective goods as a function of group size. *American Political Science Review, 68*(2), 707-716.

Chompalov, I., Genuth, J., & Shrum, W. (2002). The organization of scientific collaborations. *Research Policy, 31*(5), 749-767.

*Cohen, J. E. (1991). Size, age and productivity of scientific and technical research groups. *Scientometrics, 20*(3), 395-416.

Croyle, R.T. (2012). Confessions of a team science funder. *Translational Behavioral Medicine, 2*(4), 531-534.

Cummings, J. N., & Haas, M. R. (2012). So many teams, so little time: Time allocation matters in geographically dispersed teams. *Journal of Organizational Behavior, 33*(3), 316-341.

*Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703-722.

*Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy, 36*(10), 1620-1634.

*Cummings, J. N., Kiesler, S., Zadeh, R., & Balakrishnan, A. (2013). Group heterogeneity increases the risks of large group size: A longitudinal study of productivity in research groups. *Psychological Science, 24*(6), 880-890.

Dahlander, L., & McFarland, D. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly, 58*(1), 69-110.

Defazio, D., Lockett, A., & Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. *Research Policy, 38*(2), 293-305.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*(3), 497-509.

Fanelli, D., & Larivière, V. (2016). Researchers' individual publication rate has not increased in a century. *PLoS ONE, 11*(3): e0149504.

Fiore, S. M. (2008). Interdisciplinarity as teamwork - How the science of teams can inform team science. *Small Group Research*, *39*(3), 251-277.

Foster, J., Rzhetsky, A. & Evans, J. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review, 80*, 875-908.

Funk, R. J., & Owen-Smith, J. (2016). A dynamic network measure of technological change. *Management Science, 63*(3), 791 - 817.

Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697-702.

Guinan, E.C., Boudreau, K.C., & Lakhani, K.R. (2013). Experiments in open innovation at Harvard Medical School. *MIT Sloan Management Review, 54*(3), 45-52.

Hackman, J. R. (1987). *The design of work teams*. In J. Lorsch (Ed.), Handbook of Organizational Behavior. Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R., & Vidmar, N. (1970). Effects of size and task type on group performance and member reactions. *Sociometry, 33*, 37-54.

Hagstrom, W. (1964). Traditional and modern forms of scientific teamwork. *Administrative Science Quarterly, 9*, 241-63.

Hall, K., Stokols, D., Stipelman, B., Vogel, A., Feng, A., Masimore, B., Morgan, G., Moser, R., Marcus, S., & Berrigan, D. (2012). Assessing the value of team science: a study comparing center- and investigator-initiated grants. *American Journal of Preventive Medicine, 42*(2), 157-63.

Hill, G. W. (1982). Group versus individual performance: Are n + 1 heads better than one? *Psychological Bulletin, 91*(3), 517-539.

Hoegl, M. & Weiss, M. (2016). Effects of relative team size on teams with innovative tasks: An understaffing theory perspective. *Organizational Psychology Review, 6*(4), 324-351.

Hoegl, M., & Proserpio, L. (2004). Team member proximity and teamwork in innovative projects. *Research Policy, 33*(8), 1153-1165.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspective, 10*,153-58.

Ingham, A., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10*, 371-384.

*Jones, B., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science, 322*(5905), 1259-1262.

Katz, J., & Martin, B. (1997). What is research collaboration? *Research Policy, 26*, 1-18.

Kelley, H. & Thibaut, J. (1954). Experimental studies of group problem solving and process. In G. Lindzey (Ed.) *Handbook of Social Psychology*. Addison-Wesley: Cambridge, MA.

Kozlowski, S. & Ilgen, D. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*(3), 77-124.

*Lariviere, V., Gingras, Y., Sugimoto, C. & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology, 66*, 1323-1332.

Larrick, R., & Soll, J. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111 - 127.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*(6), 822-832.

Leahey, E., Beckman, C., & Stanko, T. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly, 62*(1), 105-139.

Leahey, E. (2016). From sole investigator to team scientist: trends in the practice and study of research collaboration. *Annual Review of Sociology, 42,* 81-100.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science, 35*, 673-702.

*Lee, Y-N, Walsh, J., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy, 44*(3), 684-697.

Lichtenstein, R., Alexander, J., Jinnett, K., & Ullman, E. (1997). Embedded intergroup relations in interdisciplinary teams: Effects on perceptions of level of team integration. *Journal of Applied Behavioral Science, 33*(4), 413-434.

Malone, T., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys, 26*(1), 87-119.

Mao, A., Mason, W., Suri, S., & Watts, D.J. (2016). An experimental study of team size and performance on a complex task. *PLOS ONE, 11*(4), e0153.

McGrath, J. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.

Merali, Z. (2010). Physics: The large human collider. *Nature, 464*, 482-484.

Milojevic, S. (2014). Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences*, *111*, 3984-3989.

Mueller, J. S. (2012). Why individuals in larger teams perform worse. *Organizational Behavior and Human Decision Processes, 117*, 111-124.

National Research Council. (2015). *Enhancing the effectiveness of team science*. In Cooke, N. Hilton, M., et al. (pp. 1-245). The National Academies Press: Washington, DC.

Nelson, P (1974). Advertising as information. *Journal of Political Economy, 82*, 729-754.

O'Leary, M.B., Mortensen, M., & Woolley, A.W. (2011). Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. *Academy of Management Review, 36*(3), 461-478.

Pendharkar, P., & Rodger, J. (2009). The relationship between software development team size and software development cost. *Communications of the ACM, 52*(1), 141-144.

Petersen, A., Pavlidis, I., & Semendeferi, I. (2014). A quantitative perspective on ethics in large team science. *Science and Engineering Ethics, 20*, 923-945.

Riedl, C., & Woolley, A. (2017). Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. *Academy of Management Discoveries, 3*(4), 382-403.

Salazar, M., Lant, T., Fiore, S., & Salas, E. (2012). Facilitating innovation in diverse science teams through integrative capacity. *Small Group Research, 43*(5), 527-558.

Seglen, P., & Aksnes, D. (2000). Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research. *Scientometrics, 49*(1), 125–143.

Seljak, U. (2012). Astrophysics: First results from Planck observatory. *Nature*, *482*, 475-477.

Shi, X., Adamic, L., Tseng, B., & Clarkson G. (2009). The impact of boundary spanning scholarly publications and patents. *PLOS ONE, 4*:e6547.

Shibayama, S., Baba, Y, & Walsh, J. (2015). Organizational design of university laboratories: Task allocation and lab performance in Japanese bioscience laboratories. *Research Policy, 44*(3), 610-622.

*Singh, J., & Fleming. L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science, 56*(1), 41-56.

Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3), 355-374.

Staats, B., Milkman, K., & Fox C. (2012). The team scaling fallacy: Underestimating the declining efficiency of larger teams. *Organizational Behavior and Human Decision Processes, 118*(2), 132-142

Stahl, M., Leap, T., & Wei, Z. (1988). Publication in leading management journals as a measure of institutional research. *Academy of Management Journal, 31*(3), 707-720.

Stallings J, et al. (2013). Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences*, *110*, 9680-9685.

Steiner, I. (1972). *Group process and productivity*. New York: Academic Press.

Stokols, D., Hall, K., Taylor, B., & Moser, R. (2008). The science of team science: Overview of the field and introduction of the supplement. *American Journal of Preventive Medicine, 35*, S77-S89.

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*, 57-74.

Thomas, E. J., & Fink, C. F. (1963). Effects of group size. *Psychological Bulletin, 60*(4), 371-384.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science, 342*(6157), 468-472.

Wagner, C., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618.

Walsh, J., & Maloney, N. (2007). Collaboration structure, communication media and problems in scientific work teams. *Journal of Computer-Mediated Communication, 12*(2), 712-732.

Wu, L., Wang, D., & Evans, J. (2018). Large teams have developed science and technology; Small teams have disrupted it. Working Paper. https://arxiv.org/abs/1709.02445

*Wuchty, S., Jones, B., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Zhang, X., Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review, 101*(4), 1601-1615.