Research Article

Psychological Science
XX(X) 1–11

DOI: 10.1177/0956797612463082
pss.sagepub.com

Group Heterogeneity Increases the Risks of Large Group Size: A Longitudinal Study of Productivity in Research Groups

Jonathon N. Cummings[1], Sara Kiesler[2], Reza Bosagh Zadeh[3], and Aruna D. Balakrishnan[2]
[1]Fuqua School of Business, Duke University; [2]Human-Computer Interaction Institute, Carnegie Mellon University; and [3]Language Technologies Institute, Carnegie Mellon University

Abstract
Heterogeneous groups are valuable, but differences among members can weaken group identification. Weak group identification may be especially problematic in larger groups, which, in contrast with smaller groups, require more attention to motivating members and coordinating their tasks. We hypothesized that as groups increase in size, productivity would decrease with greater heterogeneity. We studied the longitudinal productivity of 549 research groups varying in disciplinary heterogeneity, institutional heterogeneity, and size. We examined their publication and citation productivity before their projects started and 5 to 9 years later. Larger groups were more productive than smaller groups, but their marginal productivity declined as their heterogeneity increased, either because their members belonged to more disciplines or to more institutions. These results provide evidence that group heterogeneity moderates the effects of group size, and they suggest that desirable diversity in groups may be better leveraged in smaller, more cohesive units.

Keywords
performance, social interaction, social structure

Received 4/22/12; Revision accepted 8/31/12

An enduring interest in the field of psychology is to understand how working in groups affects member and group productivity (Hackman, 2002; Levine & Moreland, 1998). Because science is increasingly performed in groups, this question applies to how scientists conduct research. Evidence of the change in science is seen in the growing number of coauthored scientific papers; the number of authors per publication rose from two in the 1960s to almost four in the 2000s (Wuchty, Jones, & Uzzi, 2007). Many research groups are large, complex arrangements of scientists from different disciplines and institutions. In the study reported here, we examined how member heterogeneity and group size affect the productivity of such groups. Although the effects of group size (Mueller, 2012; Wheelan, 2009) and group heterogeneity (Mannix & Neale, 2005; Williams & O'Reilly, 1998) have been addressed in different streams of research, they have not been linked theoretically, particularly in the context of research groups in science and other fields. We conducted a longitudinal study of productivity in research groups to ask the following question: How is research productivity related to group heterogeneity and group size? Our analysis suggests that group heterogeneity is particularly challenging in large research groups.

Group Heterogeneity

Lab and field studies suggest that group heterogeneity derived from member differences in knowledge, expertise, or experience can increase group creativity, but only if group members build on their social and intellectual differences and work on behalf of the group as a whole (Ancona & Caldwell, 1992; Homan, van Knippenberg,

Corresponding Author:
Sara Kiesler, Carnegie Mellon University, Human-Computer Interaction Institute, 5000 Forbes Ave., Pittsburgh, PA 15232
E-mail: kiesler@cs.cmu.edu

Van Kleef, & De Dreu, 2007). This goal can be difficult to achieve. Group heterogeneity creates barriers to identification with the group as a whole because members do not feel psychologically connected to those who are different (O'Reilly, Caldwell, & Barnett, 1989; Tsui, Egan, & O'Reilly, 1992). Social identity theory (e.g., Abrams & Hogg, 1990; Brewer, 1991; Tajfel & Turner, 1986) suggests why it is particularly difficult for members to overcome differences with other members and identify with the group as a whole. Because people define themselves in terms of their meaningful social groups, they tend to view in-group members more favorably than out-group members. Categorizing people as having a different identity than oneself can lead to stereotyping and prejudice; thus, it is not surprising that members of heterogeneous groups enjoy their interactions less and like each other less well than do members of homogeneous groups (Mannix & Neale, 2005; Williams & O'Reilly, 1998).

Heterogeneity in Large Groups

The classic view of group size in the social-psychology literature is Steiner's (1972) theory of group productivity. According to Steiner, having more members provides more resources to meet task demands. Larger groups sometimes perform better than smaller groups as a result of having more people (e.g., when recalling a piece of information). Nonetheless, the potential productivity gained from having more people working on parts of the task can be offset by process losses associated with the need to motivate members to participate and coordinate their work. In larger groups, each member contributes less, on average, than in smaller groups (Liden, Wayne, Jaworski, & Bennett, 2004). One reason for this decline in marginal productivity is social loafing; members of larger groups perform less than their share of the work (Latané, Williams, & Harkins, 1979). Also, larger groups have more difficulty than smaller groups reaching a common definition of the group's goals, managing the flow of work, sustaining members' attention and cooperation, minimizing turnover, and encouraging knowledge sharing over time (Chompalov, Genuth, & Shrum, 2002; S. E. Jackson et al., 1991; Malone, 1987; Okhuysena & Bechky, 2009).

Greater heterogeneity, and thus weakened identification with the group as a whole, should exacerbate motivation and coordination costs in larger groups. Motivation costs will rise because larger groups that are heterogeneous have to spend extra effort managing and sustaining positive member relationships (Mueller, 2012). Psychological distance is greater in larger than in smaller groups (Latané, 1981); members will have less motivation to overcome and build on their differences. Meetings are less spontaneous, more formal, and less interactive in large groups (Fay, Garrod, & Carletta, 2000), which makes it harder to form bonds through informal communication. Group members are then more likely to remain more identified with their smaller (and more homogeneous) subgroup than with the larger heterogeneous group (Hinds & Mortensen, 2005). Coordination costs will rise as well because larger groups must coordinate many people's work (Mueller, 2012), and heterogeneity will increase the differences among different approaches to work. In short, as a result of the additional motivation and coordination burdens of larger groups, weakened group identification resulting from heterogeneity should undermine productivity more in these groups than in smaller groups.

In most field studies of group size and heterogeneity, researchers have examined either group heterogeneity (Mannix & Neale, 2005; Mueller, 2011) or group size (Haleblian & Finkelstein, 1993; S. E. Jackson et al., 1991) as control variables or main effects. Further, group productivity is often measured subjectively (Cummings & Kiesler, 2005, 2007). In the current study, we were interested in how group heterogeneity moderates the effects of group size on group productivity measured objectively and over time. We propose, following Steiner (1972), that group productivity will be higher in larger groups because more people contribute to the whole, but we draw on social identity theory to argue that these performance improvements will be marginally reduced as group heterogeneity increases.

Size and Heterogeneity of Research Groups

A belief that scientists gain from exposure to different approaches, and that important problems require heterogeneous research groups, has taken hold across the sciences. Rather than depending on the gradual flow of ideas from one field to another, policymakers are promoting research that integrates the contributions of different experts no matter where they reside (e.g., Cacioppo, 2007). Funding agencies, such as the National Institutes of Health, the National Science Foundation (NSF), and the European Union's Framework Programme, have promoted team science and funded sizeable projects that span disciplines and universities (Finholt & Olson, 1997). These projects are tackling complex topics in areas such as neuroscience, bioengineering, and medicine (Corley, Boardman, & Bozeman, 2006; Jordan, 2006; Metzger & Zare, 1999).

Our theoretical analysis suggests that increasing the heterogeneity and size of research groups will exact additional process losses in those groups. For instance, it may be essential for a research group's goals to recruit and add experts from another discipline, but these experts are unlikely to share the same social identity as the rest of the group. The group must then expend extra effort to develop trust and overcome differences of language and norms about the research process (Palmer, 1999). The same can be said of a research group that is dispersed and comprised of people from different institutions (Herbsleb, Mockus, Finholt, & Grinter, 2000; Olson & Olson, 2000). The group, at the same time, typically has become larger, so it faces bigger motivation and coordination challenges. Without strong group identification, these challenges may not be met. For instance, a small group of researchers can use conventional online tools to talk, share resources, arrange meetings, come to know each other's students, and learn about each other's perspectives and skills, all of which should contribute to their identification with the group. In a large research group, however, addressing heterogeneity is more difficult. As anyone who belongs to a large research group can attest, finding meeting times for a large group, much less carrying out informal collegial communication, can be difficult.

We propose that heterogeneity interferes with the marginal productivity of large research groups by reducing the additional gains that ordinarily would accrue as groups get larger. In one longitudinal analysis of awarded grants, the director of the National Institute of General Medical Sciences reported that the largest labs were significantly less productive and had less impact than mid-size labs (Wadman, 2010), but many other factors that could have reduced large labs' productivity, including their heterogeneity, were not controlled in this study. We therefore set out to examine the hypothesis that group heterogeneity moderates the effects of group size on group research productivity.

A Longitudinal Study of the Effects of Group Size and Heterogeneity in Research Groups

Method

We examined the relationships of group size, heterogeneity, and group productivity in 549 research groups. These groups were created to carry out Information Technology Research (ITR) projects funded by NSF from 2000 to 2004 with up to $3 million per year for 5 years. ITR was a 5-year NSF-wide priority area that grew from $90 million in 2000 to $295 million in 2004. Together, these projects involved more than 2,200 principal investigators (PIs), along with numerous other researchers, staff, and students. A typical research group comprised five PIs and their students from two top U.S. universities, had representation from two or three disciplines, and received $2 million in funding over 5 years (Table 1). Many projects were aimed at developing techniques or theory from computer science for other disciplines, such as biology, physics, engineering, and psychology. The program was very popular (2,100 proposals were submitted in the first year of the program), and the program became more competitive over time. In 2000, 30% of the medium and large proposals were funded with 70% of their proposed budget, but by 2004, just 21% were funded with 49% of their budget.

Productivity measures. To assess these projects' outcomes, in 2009, 4 to 9 years after the projects started, we used four measures of productivity. One measure was the number of publications the PIs listed in their final reports to NSF (or in the latest annual report, if no final report had been submitted by the time of our study). Publications in final reports include archival conference proceedings, journal articles, chapters in books, and public reports on the project. When considering project publications, we pooled all publications authored by PIs and removed duplicate publications when PIs coauthored a paper.

The three additional productivity measures were group member's cumulative publications obtained using the Google Scholar search engine, and their cumulative publications and citations in the Thomson Reuters (formerly ISI) Web of Science and Social Science database. We divided their publications into those published prior to the start date of the ITR project and those published after this date. To check the quality of these automatically extracted publications, we evaluated 10% of the sample manually using participants recruited via Amazon's Mechanical Turk. For each extracted publication, we asked participants to find the corresponding author's Web page or résumé with their publications listed and check that the automatically extracted publication was indeed correct. We asked 5 participants to check each publication and assumed correctness if at least 4 gave the same result. Overall, 94% of the extracted publications were correct.

Group size and heterogeneity. We measured the size of research groups by counting the total number of PIs listed on the project grant. This information was obtained from the awards database posted on the NSF (2009) Web site. We measured heterogeneity in two ways: (a) by the number of PIs' disciplines and (b) by the number of institutions in which the PIs worked. We obtained the disciplines of PIs from their departmental affiliation and manually checked those that were ambiguous. Their institutions were listed in the NSF awards Web site.

Control variables. A number of factors other than group size and heterogeneity can influence the productivity of research groups. As noted previously, we used

Table 1. Means, Standard Deviations, and Correlations Among the Variables in the Study

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
Control variables															
1. Project-year start	2002.3	1.3	1.00												
2. R&D expenditures of institutions[a]	4.2	0.7	−.07	1.00											
3. Project budget	$1.9M	$2.2M	−.19	.05	1.00										
4. Number of active NSF grants	2.9	2.7	−.03	.01	.25	1.00									
5. Google Scholar publications before project[b]	154.8	95.5	.02	−.02	.49	.47	1.00								
6. Web of Science publications before project[b]	78.9	68.2	.28	−.03	.36	.33	.71	1.00							
7. Web of Science citations before project[b]	920.9	1,129.9	.21	−.02	.29	.24	.53	.77	1.00						
Group size and heterogeneity															
8. Number of PIs	4.9	3.0	.00	−.07	.49	.47	.94	.73	.55	1.00					
9. Number of institutions of PIs	2.3	1.6	−.01	−.13	.37	.37	.67	.52	.46	.72	1.00				
10. Number of disciplines of PIs	2.1	0.9	.02	−.01	.18	.25	.53	.40	.32	.57	.33	1.00			
Productivity outcomes															
11. Google Scholar publications[c]	151.9	95.6	−.09	−.01	.51	.48	.94	.66	.48	.92	.65	.53	1.00		
12. Web of Science publications[c]	151.8	102.6	−.04	−.06	.48	.47	.90	.72	.57	.93	.65	.55	.90	1.00	
13. Web of Science citations[c]	1,028.8	1,240.2	−.20	.05	.48	.25	.54	.43	.48	.54	.37	.31	.54	.61	1.00
14. Final-report publications[d]	85.5	67.6	.09	−.07	.36	.25	.55	.40	.30	.54	.35	.33	.57	.51	.30

Note: N = 549. M = million; NSF = National Science Foundation; PI = principal investigator.
[a]Institutions were scored from 1, *bottom 20% of institutions*, to 5, *top 20% of institutions*, according to their level of research and development (R&D) expenditures. [b]These variables index the number of unique publications or citations of PIs prior to the start of the current project. [c]These variables index the number of unique publications or citations of PIs from the start of their project to 2009. [d]This variable represents the number of publications reported in the project's final report (or the last annual report if the project was ongoing).

measures of each PI's publications (or citations) prior to their ITR project to control for their research productivity at the start of the project. We controlled for the number of other active NSF grants held by the PIs at project start because PIs' other grants provide additional resources. We also controlled for the start date of the project because older projects will have had more time to work. Because ITR was a new program, we also entered a quadratic factor to account for projects funded in the first year that could have had a particularly rough start. We also controlled for the average research and development funding of the universities involved in the project because universities whose faculty has more research experience may provide better institutional support. Jones, Wuchty, and Uzzi (2008) show that multiinstitutional collaborations are increasingly concentrated within top-tier universities and that coauthored papers are more highly cited when they include an author from a top-tier university. We also controlled for the amount of funding awarded to the project.

Analyses. We examined all projects' publication output from their start date (2000–2004) through 2009, which yielded 46,850 publications listed in project reports to NSF. We then used hierarchical regression models to examine the predictors of these publications, controlling statistically for PIs' productivity before their project began (an estimate of the baseline likelihood of their future productivity). An important aspect of this prospective methodology is that we captured considerable variation in number of disciplines, number of institutions, and group size, as well as variation in the marginal productivity of projects. We were not limited to successful projects only or to those PIs who had published papers.

In our regression analysis, we tested the hypothesis by assessing how groups comprised of PIs from many disciplines or universities fared when the groups varied in size. To remove the possibly undue effect of unusually large or heterogeneous projects, we used truncated measures of group size (2–13+ investigators), disciplinary heterogeneity (1–4+), and institutional heterogeneity (1–7+) in all analyses. In the statistical analyses, when determining the quantity of publications—whether from NSF final reports, Google Scholar publications, or Web of Science publications—we counted a group's publication only once, no matter how many authors it had from the project. Similarly, in determining the quality or impact of publications using Web of Science citations, we relied on the citations for each unique publication.

Interviews. To obtain information on PIs' retrospective perceptions of motivation and coordination in the research groups, in 2009, we conducted structured interviews with 55 PIs from 52 of the research projects. The sample was chosen blind to the productivity data. We drew a stratified random sample from projects at top-ranked universities in locations reflecting the distribution of projects overall: 15 researchers from the Northeast, 13 from the South, 7 from the Midwest, and 20 from the West. When possible, we interviewed the lead PI. Our interview questions drew from Kraut, Galegher, and Egido's (1987) model of research collaboration, which posits three stages of research: the initiation phase (e.g., preparing a proposal, finding experts), the execution phase (e.g., collecting data, running experiments), and the dissemination phase (e.g., writing and publishing papers). PIs were encouraged to discuss their project experiences, how they chose their collaborators, how they planned their budgets and projects, and how they organized their work and wrote papers. We coded interviews iteratively using NVivo software (Version 10; QSR International, Doncaster, Victoria, Australia) and then used frequently coded themes to help develop our theoretical arguments and derive plausible explanations of the longitudinal study results.

Results

Table 1 shows the means and correlations among the variables. We consider the publications listed in the PIs' final reports to be the best measure of productivity in research groups, because these reports would be less likely than Google Scholar or Web of Science to include the results of PIs' other projects and collaborations. (PIs had to acknowledge the grant in the publications they listed.)

Longitudinal analyses. Tables 2 through 5 present the results of the hierarchical regression analyses using four dependent measures of group productivity, controlling for prior PI research productivity. As the tables show, group size had a significantly positive effect on productivity. This result was expected because as each PI is added, there is one more researcher to contribute to the group's publishable research. We also found that heterogeneity did not generally influence productivity (although more institutions in a group was a consistently negative trend).

As hypothesized, the interactions of group size and heterogeneity were statistically significant (Group Size × Number of Disciplines: $\beta = -0.13$, $p < .01$; Group Size × Number of Institutions: $\beta = -0.13$, $p < .01$). These interactions are displayed in Figure 1. Having more PIs predicted higher productivity overall, but when the research group was heterogeneous, either because the PIs in it had multiple disciplines or because the PIs came from multiple institutions, the results changed. Simple-slopes analyses (Aiken & West, 1991) confirmed that productivity in

Table 2. Results From Hierarchical Regression Models Predicting the Effect of Group Size and Heterogeneity on Group Research Productivity, as Measured by Log NSF Final Report Publications

Predictor	Step 1	Step 2	Step 3	Step 4
Control variables				
Log publications prior to project[a]	0.33***	0.17***	0.12*	0.10†
Number of active NSF grants at time of project start	0.02	–0.09	0.00	0.00
Project-year start	–0.02	0.04	–0.03	–0.03
Project Year × Project Year	–0.47***	–0.47***	–0.47***	–0.47***
Average R&D funding of project institutions	–0.06†	–0.04	–0.03	–0.03
Log (project funding)	0.20***	0.16***	0.16***	0.16***
Key predictors				
Number of investigators (2–13+)[b]	—	0.27***	0.37***	0.37***
Number of disciplines (1–4+)[b]	—	0.02	0.00	–0.03
Number of institutions (1–7+)[b]	—	–0.07	–0.04	–0.06
Number of Investigators × Number of Disciplines	—	—	–0.11*	–0.13**
Number of Investigators × Number of Institutions	—	—	–0.10*	–0.13**
Number of Disciplines × Number of Institutions	—	—	0.08	0.03
Number of Institutions × Number of Investigators × Number of Disciplines	—	—	—	0.14*
R^2 (adjusted)	.41	.43	.44	.44

Note: N = 549. Standardized coefficients are shown. NSF = National Science Foundation; R&D = research and development.
[a]This variable was estimated from Google Scholar cumulative publications prior to the project start. [b]These independent variables were truncated.
†$p < .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Table 3. Results From Hierarchical Regression Models Predicting the Effect of Group Size and Heterogeneity on Group Research Productivity, as Measured by Log Google Scholar Publications

Predictor	Step 1	Step 2	Step 3	Step 4
Control variables				
Log publications prior to project[a]	0.82***	0.65***	0.62***	0.61***
Number of active NSF grants at time of project start	0.07**	0.04†	0.04*	0.04†
Project-year start	–0.15***	–0.16***	–0.16***	–0.16***
Project Year × Project Year	–0.05*	–0.06**	–0.05**	–0.06**
Average R&D funding of project institutions	–0.02	–0.00	0.00	0.01
Log (project funding)	0.07***	0.03	0.04†	0.04†
Key predictors				
Number of investigators (1–13+)[b]	—	0.21***	0.27***	0.27***
Number of disciplines (1–4+)[b]	—	0.04	0.03	0.02
Number of institutions (1–7+)[b]	—	–0.01	0.00	–0.00
Number of Investigators × Number of Disciplines	—	—	–0.07**	–0.09**
Number of Investigators × Number of Institutions	—	—	–0.06†	–0.07*
Number of Disciplines × Number of Institutions	—	—	–0.03	–0.00
Number of Institutions × Number of Investigators × Number of Disciplines	—	—	—	0.09*
R^2 (adjusted)	.78	.79	.80	.80

Note: N = 549. Standardized coefficients are shown. NSF = National Science Foundation; R&D = research and development.
[a]This variable was estimated from Google Scholar cumulative publications prior to the project start. [b]These independent variables were truncated.
†$p < .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Table 4. Results From Hierarchical Regression Models Predicting the Effect of Group Size and Heterogeneity on Group Research Productivity, as Measured by Log Web of Science Publications

Predictor	Step 1	Step 2	Step 3	Step 4
Control variables				
Log publications prior to project[a]	0.65***	0.45***	0.39***	0.33***
Number of active NSF grants at time of project start	0.18***	0.07**	0.08**	0.07**
Project-year start	−0.21***	−0.17***	−0.15***	−0.14***
Project Year × Project Year	−0.01	0.01	0.01	—
Average R&D funding of project institutions	−0.05†	−0.02	−0.01	−0.00
Log (project funding)	0.08**	−0.02	−0.01	−0.01
Key predictors				
Number of investigators (1–13+)[b]	—	0.41***	0.53***	0.53***
Number of disciplines (1–4+)[b]	—	0.08*	0.07*	0.03
Number of institutions (1–7+)[b]	—	−0.06†	−0.02	−0.04
Number of Investigators × Number of Disciplines	—	—	−0.19***	−0.22***
Number of Investigators × Number of Institutions	—	—	−0.13***	−0.17***
Number of Disciplines × Number of Institutions	—	—	0.09*	0.00
Number of Institutions × Number of Investigators × Number of Disciplines	—	—	—	0.20***
R^2 (adjusted)	.54	.63	.67	.67

Note: N = 549. Standardized coefficients are shown. NSF = National Science Foundation; R&D = research and development.
[a]This variable was estimated from Web of Science cumulative publications prior to the project start. [b]These independent variables were truncated.
†$p < .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Table 5. Results From Hierarchical Regression Models Predicting the Effect of Group Size and Heterogeneity on Group Research Productivity, as Measured by Log Web of Science Citations

Predictor	Step 1	Step 2	Step 3	Step 4
Control variables				
Log citations prior to project[a]	0.62***	0.52***	0.48***	0.47***
Number of active NSF grants at time of project start	0.11***	0.03	0.04	0.04
Project-year start	−0.26***	−0.25***	−0.24***	−0.24***
Project Year × Project Year	−0.06†	−0.05	−0.05	−0.05
Average R&D funding of project institutions	0.02	0.02	0.03	0.03
Log project funding	0.10**	0.03	0.04	0.04
Key predictors				
Number of investigators (1–13+)[b]	—	0.31***	0.39***	0.38***
Number of disciplines (1–4+)[b]	—	0.03	0.02	0.00
Number of institutions (1–7+)[b]	—	−0.11**	−0.08†	0.09*
Number of Investigators × Number of Disciplines	—	—	−0.14**	−0.16***
Number of Investigators × Number of Institutions	—	—	−0.10*	−0.11**
Number of Disciplines × Number of Institutions	—	—	0.08†	0.04
Number of Institutions × Number of Investigators × Number of Disciplines	—	—	—	0.09
R^2 (adjusted)	.47	.51	.53	.53

Note: N = 549. Standardized coefficients are shown. NSF = National Science Foundation; R&D = research and development.
[a]This variable was estimated from Web of Science citations prior to the project start. [b]These independent variables were truncated.
†$p < .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

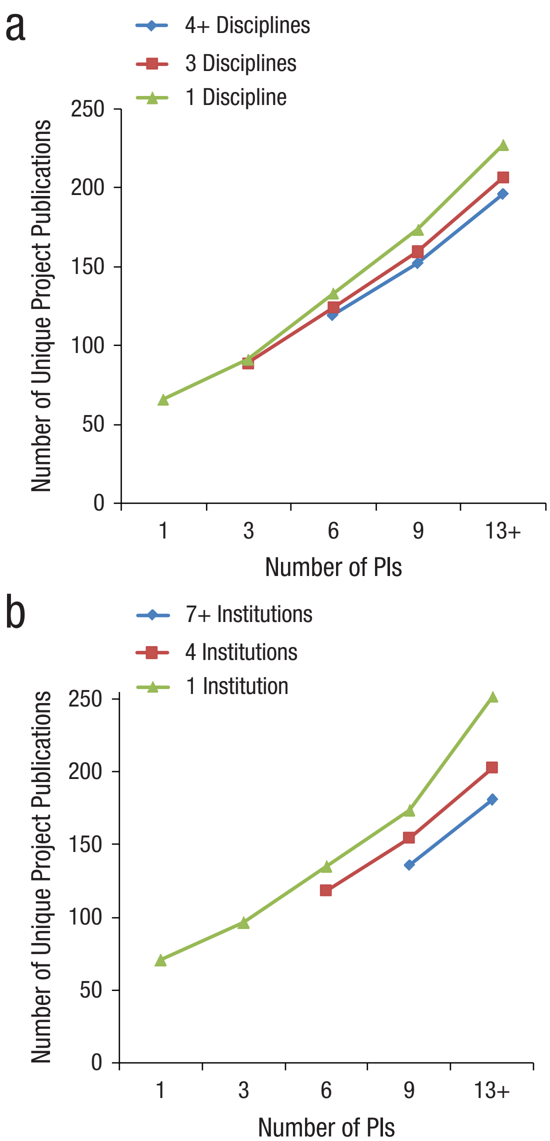

Fig. 1. Predicted number of unique project publications as a function of group size and heterogeneity in research groups. Group heterogeneity was measured by either (a) the number of disciplines of the principal investigators (PIs) in a group or (b) the number of institutions involved in the research.

research groups lowest in heterogeneity (one discipline or one institution) increased with more members—one discipline: $t(1) = 5.23$, $p < .0001$, $d = 0.45$; one institution: $t(1) = 4.88$, $p < .0001$, $d = 0.42$. At medium levels of heterogeneity (three disciplines or four institutions), productivity in groups also increased with more members, but not as much—three disciplines: $t(1) = 2.79$, $p < .01$, $d = 0.24$; four institutions: $t(1) = 2.5$, $p = .01$, $d = 0.22$. Productivity in groups highest in heterogeneity (four or more disciplines or seven or more institutions) did not increase with more group members—four or more disciplines: $t(1) = 0.64$, n.s.; seven or more institutions: $t(1) = 0.13$, n.s. (see Figs. 1a and 1b).

To get a sense of these effects, consider that the average project produced 85.5 unique publications (not double-counting project members on the same publication). Controlling for other factors, we found that a typical five-member group (that is, mean and median size) whose members came from three disciplines produced 119 publications, whereas a larger group of nine PIs whose members came from three disciplines produced 150 publications. Although 150 is greater than 119, the output per PI is nearly 24 publications in the smaller group and just 17 publications in the larger group. In a group as large as 13, per-PI output went down to 14 publications. The pattern and significance of the moderator effect was the same when we used different measures of productivity, including citations. Tables 2 through 5 show the regressions predicting the interaction effects of size and the number of disciplines on productivity—Google Scholar publications: $\beta = -0.09$, $p < .01$; Web of Science publications: $\beta = -0.22$, $p < .001$; Web of Science citations, $\beta = -0.16$, $p < .001$—and the effects of size and more institutions involved in the project—Google Scholar publications: $\beta = -0.07$, $p < .05$; Web of Science publications: $\beta = 0.17$, $p < .001$; Web of Science citations, $\beta = -0.11$, $p < .001$.

Interview findings. Among the 55 interviewees, many recalled problems in communication that they attributed to heterogeneity (64%) or to large group size (34%). Some said these problems had interfered with sharing information (55%) or resources (18%) and had led PIs to go their separate ways (24%). For example, when members were writing their proposals, some groups added PIs from another discipline to obtain more expertise on the topic and to bolster their interdisciplinary credentials, but lack of familiarity interfered with group chemistry and encouraged members to work with group members they already knew. (For example, one interviewee said, "Why would I want to build a personal relationship and start work with someone else when I could work with my buddy? It's more fun.") When executing research, PIs may have intended to use project resources to support groundbreaking interdisciplinary work, but their first responsibility, in their view, was to their own part of the project, especially to helping their students publish in the top journals of their discipline. At the dissemination stage, the process of reporting interdisciplinary work in specific venues of interest to each discipline strained relationships. These findings suggested to us that a lack of identification and integration with the research project as a whole was a key failing of large, heterogeneous groups.

Discussion

Our analysis of the productivity of 549 research groups indicated that more PIs on a project increased productivity, but heterogeneity reduced the marginal productivity of research groups when members were from multiple disciplines or institutions. These results are not due to the personal productivity of PIs, their access to other funding, the duration of their projects, differences in their project budgets, or their universities' experience with research. Both forms of heterogeneity we measured, multiple disciplines and multiple universities, were problematic.

Working across disciplines entails different communication challenges than working across institutions. Yet both of these situations require that PIs actively manage relationships and accommodate different perspectives. Our findings suggest that most researchers struggled to perform these tasks in large projects. Reviewing our data, however, we identified a few heterogeneous large research groups with unusually high publication rates. An interview we had with the lead PI of one of these groups suggested that having a strong leader who insists on frequent project meetings and status reports from everyone might help large, heterogeneous groups overcome member differences. This PI said, "One of the advantages [was that] I was PI. And I have worked in this cross-disciplinary space for a long time. And so basically people knew I wouldn't tolerate any hiding in your discipline. So it was like if you're not part of this cultural change to meld together across these things then we don't need you on the project." Another interview with the PI of a successful large group suggested that having a balance of expertise at each site or in each discipline rather than token or unconnected experts could help people cross subgroup boundaries (cf. J. W. Jackson, 1999): "An awful lot of the work is learning to understand each others' vocabulary. . . I don't know a lot about her field and vice versa. . . It helped that [in my lab] I had a junior faculty member [in the other field] working on the project as well and so he could act as the translator between the two of us." We believe that such "translation" activities might have helped the members develop common goals and stronger group identity.

Conclusion

Diversity of perspectives and skills in a group makes innovation possible, but acquiring this diversity may mean adding members. Our data suggest that there are limits to the advantages of adding people and that diversity may be applied better in smaller, more manageable groups than in larger groups. Examining other group tasks will be important to test the generalizability of our findings. For example, in creative-design groups, the benefits of heterogeneity may outweigh the costs of having more members. Steiner (1972) suggested that large groups perform better if they can easily subdivide tasks and reassemble the results. This argument foreshadows the advent of crowd-sourced science, whereby ordinary citizens and scientists collect or analyze data for the benefit of the whole. We also do not know whether our findings generalize to really large research groups on the order of what has been required to sequence the human genome or conduct high-energy physics experiments. Perhaps a 15-person research group is very different than a 100- to 1,000-person research organization. We also do not know the time frame in which heterogeneity begins to cause trouble for larger groups or the extent to which groups can overcome their differences. Although heterogeneous groups may just take longer to become productive when they are large, our sense from the interviews is that if there are few successes early on, they are unlikely to come later.

The impact of perceiving the entire group as a cohesive unit has been studied as a problem in entitativity (Lickel et al., 2000). It would seem plausible that if members identified mainly with those in their own discipline or with others at their own institution, then the research group could be considered a collection of subgroups (Carton & Cummings, 2012). If so, psychological distance among subgroups might be an important theoretical mechanism linking heterogeneity, group size, and productivity (see Trope & Liberman, 2010).

Within the limitations of the data from the NSF program we studied, we have shown that large, diverse groups are not as marginally productive as comparatively smaller diverse groups. We hope these findings encourage more research on the processes of managing heterogeneity and group size. Given today's complex problems, we need better ways to more constructively marshal a variety of people and resources to tackle them.

Acknowledgments

We thank Charles Ball and Ian Shields for their excellent research assistance, and Suzanne Iacono and Wesley Cohen for their helpful comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by collaborative research awards from the National Science Foundation (OCI-0838367 and SBE-0830254 to Duke University, and OCI-0838385 and SBE-0830306 to Carnegie Mellon University).

References

Abrams, D., & Hogg, M. A. (1990). *Social identifications: A social psychology of intergroup relations and group processes*. London, England: Routledge.

Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Ancona, D., & Caldwell, D. (1992). Demography & design: Predictors of new product team performance. *Organization Science*, *3*, 321–341.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*, 475–482.

Cacioppo, J. (2007). Better interdisciplinary research through psychological science. *The Observer*, *20*(10), 3–5.

Carton, A. M., & Cummings, J. N. (2012). A theory of subgroups in work teams. *Academy of Management Review*, *37*, 441–470.

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

Corley, E. A., Boardman, P. C., & Bozeman, B. (2006). Design and the management of multi-institutional research collaborations: Theoretical implications from two case studies. *Research Policy*, *35*, 975–993.

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

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

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue. *Psychological Science*, *11*, 481–486.

Finholt, T., & Olson, G. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, *8*, 28–36.

Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston, MA: Harvard Business School.

Haleblian, J., & Finkelstein, S. (1993). Top management team size, CEO dominance, and firm performance: The moderating roles of environmental turbulence and discretion. *Academy of Management Journal*, *36*, 844–863.

Herbsleb, J. D., Mockus, A., Finholt, T., & Grinter, R. E. (2000). Distance, dependencies, and delay in a global collaboration. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work '00* (pp. 319–328). New York, NY: ACM Press.

Hinds, P., & Mortensen, M. (2005). Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science*, *16*, 290–307.

Homan, A. C., van Knippenberg, D., Van Kleef, G. A., & De Dreu, C. K. W. (2007). Bridging faultlines by valuing diversity: Diversity beliefs, information elaboration, and performance in diverse work groups. *Journal of Applied Psychology*, *92*, 1189–1199.

Jackson, J. W. (1999). How variations in social structure affect different types of intergroup bias and differential dimensions of social identity in a multi-intergroup setting. *Group Processes & Intergroup Relations*, *2*, 145–173.

Jackson, S. E., Brett, J. F., Sessa, V. I., Cooper, D. M., Julin, J. A., & Peyronnin, K. (1991). Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment, promotions, and turnover. *Journal of Applied Psychology*, *76*, 675–689.

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

Jordan, G. B. (2006). Factors influencing advances in basic and applied research: Variation due to diversity in research profiles. In J. Hage & M. Meeus (Eds.), *Innovation, science, and institutional change* (pp. 173–195). New York, NY: Oxford University Press.

Kraut, R. E., Galegher, J., & Egido, C. (1987). Relationships and tasks in scientific research collaboration. *Human-Computer Interaction*, *3*, 31–58.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*, 343–365.

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

Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 415–469). New York, NY: McGraw-Hill.

Lickel, B., Hamilton, D., Wieczorkowska, G., Lewis, A., Sherman, S., & Uhles, A. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*, 223–246.

Liden, R., Wayne, S., Jaworski, R., & Bennett, N. (2004). Social loafing: A field of investigation. *Journal of Management*, *30*, 285–304.

Malone, T. W. (1987). Modeling coordination in organizations and markets. *Management Science*, *33*, 1317–1332.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, *6*, 31–55.

Metzger, N., & Zare, R. (1999). Interdisciplinary research: From belief to reality. *Science*, *283*, 642–643.

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

National Science Foundation. (2009). *Academic R&D expenditures*. Retrieved from http://www.nsf.gov/statistics/pubseri.cfm?seri_id=19

Okhuysena, G. A., & Bechky, B. A. (2009). Coordination in organizations: An integrative perspective. *The Academy of Management Annals*, *3*, 463–502.

Olson, G. M., & Olson, J. (2000). Distance matters. *Human-Computer Interaction*, *15*, 139–179.

O'Reilly, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly*, *34*, 21–37.

Palmer, C. (1999). Structures and strategies of interdisciplinary science. *Journal of the Association for Information Systems*, *50*, 242–253.

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

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*, 440–463.

Tsui, A., Egan, T., & O'Reilly, C. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly*, *37*, 549–579.

Wadman, M. (2010). Study says middle sized labs do best. *Nature*, *468*, 356–357.

Wheelan, S. A. (2009). Group size, group development, and group productivity. *Small Group Research*, *40*, 247–262.

Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, *20*, 77–140.

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