The Team Scaling Fallacy:

Underestimating The Declining Efficiency of Larger Teams

Bradley R. Staats,[1] Katherine L. Milkman,[2] and Craig R. Fox[3]

[1]University of North Carolina at Chapel Hill

[2]The Wharton School, The University of Pennsylvania

[3]UCLA Anderson School and Department of Psychology

Forthcoming, *Organizational Behavior and Human Decision Processes*

Abstract

The competitive survival of many organizations depends on delivering projects on time and on budget. These firms face decisions concerning how to scale the size of work teams. Larger teams can usually complete tasks more quickly, but the advantages associated with adding workers are often accompanied by various disadvantages (such as the increased burden of coordinating efforts). We note several reasons why managers may focus on process gains when they envision the consequences of making a team larger, and why they may underestimate or underweight process losses. We document a phenomenon that we term *the team scaling fallacy*—as team size increases, people increasingly underestimate the number of labor hours required to complete projects. Using data from two laboratory experiments, and archival data from projects executed at a software company, we find persistent evidence of the team scaling fallacy and explore a reason for its occurrence.

Keywords: Coordination Neglect, Estimation, Planning Fallacy, Team Scaling Fallacy, Team size

The Team Scaling Fallacy: Underestimating The Declining Efficiency of Larger Teams

Across a wide range of industries and functions, from construction to consulting and from healthcare to new product development, work is delivered to customers in the form of projects completed by teams (Edmondson & Nembhard, 2009; Ilgen, Hollenbeck, Johnson, & Jundt, 2005). Organizations turn to teams for many reasons, one of which is the increased speed with which projects can be completed when work is divided among many people. Organizations also rely increasingly on teams because knowledge is evolving so rapidly that in many settings, no single person has the depth of knowledge required to adequately serve customer needs. Teams also allow for specialization of member roles through the division of labor and can increase the knowledge resources available both within a team and through members' external connections (Haleblian & Finkelstein, 1993; Moreland, Levine, & Wingert, 1996; Reagans & Zuckerman, 2001).

In many project-based organizations that rely on teams, an important key to competitive success is accurately estimating and adhering to project budgets and deadlines. For a business that delivers projects to customers, missing promised budget and deadline estimates can tarnish a previously good reputation with patrons, resulting in lost business. Such errors in forecasting may also turn projects that should have generated profits into money-losing ventures (Heskett, Sasser, & Schlesinger, 1997; Wheelwright & Clark, 1992). Despite the importance of meeting deadlines and correctly estimating costs, industry statistics suggest that many project-based organizations struggle with these activities. For example, studies in the construction, healthcare, aerospace, and information technology industries have found that anywhere from 33% to 88% of projects are delivered late and over budget (Knight, 2011; Standish, 2009; Watson, 2008).

One possible explanation for these budget and deadline overruns is that process challenges arise when people work together, yet estimators do not properly account for them. Research on teams has shown that although increasing a team's size provides the *potential* for many benefits (e.g., through increased specialization and expanded knowledge networks), the team's *actual* productivity may suffer due to process losses (Levine & Moreland, 1998; Steiner, 1972). Increasing a team's size can hamper its coordination, diminish its members' motivation, and increase conflict among team members (Hare, 1952; Ingham, Levinger, Graves, & Peckham, 1974; Moreland et al., 1996). An interesting question is whether estimators are sufficiently sensitive to these problems. In this paper, we investigate whether estimators exhibit a bias that we term the *team scaling fallacy*—a tendency to increasingly underestimate task completion time as team size grows. We confirm the hypothesis that the team scaling fallacy plagues estimators in both the laboratory and the field. We also identify and test an important driver of this phenomenon: the tendency to focus too much on the process gains associated with increasing team size, relative to the process losses.

Background and Hypotheses

Impact of Team Size on Team Performance

Before investigating the impact of team scaling on forecasting errors, it is important to consider how team size affects team performance (see Levine & Moreland, 1998 and Moreland, et al., 1996, for reviews of this topic). Increasing size offers a team multiple benefits. Labor can be subdivided across more team members, for example. This division of labor makes it possible to match workers with the tasks that are most interesting to them and for which they are best suited. It can also foster task specialization, which improves performance (Moreland & Myaskovsky, 2000; Newell & Rosenbloom, 1981; Wegner, 1987). Moreover, larger teams are likely to have a broader base of knowledge and experience (Haleblian & Finkelstein, 1993; Reagans & Zuckerman, 2001), which can prove beneficial. Larger teams also possess more slack resources, which can be deployed if circumstances change (Moreland et al., 1996).

Along with these benefits, increasing team size presents challenges involving coordination, motivation, and conflict (Hackman, 2002; Levine & Moreland, 1998; Steiner, 1972). With respect to coordination, the potential for coordination losses increases as a team grows because the number of communication linkages among members increases at a nonlinear rate.[1] More time is thus required to keep all members informed (Brooks, 1975; Stasser & Taylor, 1991). The threat of miscommunication also increases when information is passed among a greater number of members, each of whom may interpret the information differently, based on his or her personal background (Allen, 1977; Bechky, 2003). Further, although a larger team creates opportunities for division of labor, completed work must be integrated at some point, requiring additional time and effort (Heath & Staudenmayer, 2000; Lawrence & Lorsch, 1967).

A second challenge associated with increasing team size involves member motivation. Members of growing teams may experience decreased motivation. For example, members of larger teams may exert less effort, due to such factors as social loafing and free riding (Albanese & Fleet, 1985; Karau & Williams, 1993; Latane, Williams, & Harkins, 1979). Also, members of larger teams may find their membership to be less satisfying (Hackman & Vidmar, 1970; Mullen, Symons, Hu, & Salas, 1989), which can weaken their commitent.

A third challenge associated with increasing team size is the increasing potential for conflict among members, which can again harm team performance (Brewer & Kramer, 1986; O'Dell, 1968). For example, members may be less willing to help one another in a larger team, or they may suppress the ideas of others in order to promote their own ideas (Diehl & Stroebe, 1987; Latané & Nida, 1981; Paulus & Yang, 2000).

Biases in Estimating the Impact of Increasing Team Size

Our focus in this paper is not on whether team performance improves or deteriorates as a team grows. Rather, we are concerned with whether people *are sufficiently sensitive* to the impact of increasing team size on the total number of hours of labor required to complete a project.

A principal reason why people may underestimate diminishing returns to increasing team size is that they may underestimate the additional time needed to coordinate team members' efforts. This error, known as *coordination neglect* (Heath & Staudenmayer, 2000), has been hypothesized to occur when estimators attend more to the gains in efficiency that can be achieved by dividing responsibility for project components among team members than they do to the time required to integrate that work. As a team grows, its opportunities for dividing labor increase, but so does the complexity of integrating completed work. Thus, if people focus primarily on the gains from dividing labor, then their estimates of the effort required to complete a project will be increasingly overoptimistic as a team grows in size. Although coordination neglect has been discussed in the academic literature, we are aware of no published empirical tests of coordination neglect and its implications.

Past research on decision making has focused on documenting biases that lead people to make inaccurate judgments across a range of domains (see e.g., Bazerman & Moore, 2009; Gilovich, Griffin & Kahneman, 2002). Early research on judgment and decision-making demonstrated that people are poor intuitive statisticians (Tversky & Kahneman, 1974), making many errors when they estimate probabilities. More recent research has demonstrated that people also exhibit a wide range of self-serving biases, such as over-optimism about their own abilities (see Moore & Healy, 2008). For example, studies of brainstorming have shown that many people believe that groups will be more productive than individuals, even though a comparable number of individuals will consistently outperform a group when it comes to brainstorming (Diehl & Stroebe, 1987). Group brainstorming may persist because people feel more satisfied working in a group and have an inflated view of their own contribution to the brainstorming process (Nijstad, Stroebe, & Lodewijkx, 2006; Paulus & Yang, 2000).

Predictors who have a stake in a project's success may also exhibit self-serving biases when estimating project completion times. Past research has shown that many people exhibit a "planning fallacy," underestimating how much of their *own* effort will be required to complete a project alone (Buehler, Griffin, & Ross, 2002; Kahneman & Tversky, 1979) or with teammates (Buehler, Messervey, & Griffin, 2005; Sanna, Parks, Chang, & Carter, 2005). This bias is said to arise from a tendency to imagine scenarios involving success, using details of the case at hand, without adequate attention to relevant base rates of performance on similar projects in the past (Buehler, Griffin, & Ross, 1994; Kahneman & Lovallo, 1993).

Although the planning fallacy might contribute to budget and deadline overruns in many organizations, we believe that it would not generally contribute to a team scaling fallacy. Why? First, the planning fallacy does not seem to afflict outsiders—it biases only an individual's predictions about his or her *own* performance (Buehler, Griffin & Ross, 1994). However, in many organizations, those who are responsible for estimating the total labor hours required for a team to complete a project are not members of the team that will carry out the assignment. Second, even when predictors do have a stake in project outcomes, it is not clear why the severity of the planning fallacy should vary for teams of different sizes when task characteristics are otherwise held constant. Indeed, to the extent that partitioning a task among the members of larger teams causes estimators to more thoroughly "unpack" the components of a project, that process should reduce the planning fallacy (Kruger & Evans, 2004).

We study whether people who must estimate the total amount of labor required to complete a team project are appropriately responsive to differences in team size. The specific metric we examined is the error in the estimated total number of hours (or minutes) of labor required to complete the project ("effort"). This metric is more sensitive than a "missed deadline" because if a project is running behind schedule, then employees can be induced to work overtime, thus meeting their calendar deadline, yet exceeding their effort budget and thereby inflating project costs. Similarly, individuals could work fewer hours per day on a project that is ahead of schedule, appearing to barely make a deadline, when in reality, considerable slack existed.

We examine two testable hypotheses regarding the accuracy of estimators' forecasts about the effort that teams will require to complete projects as a function of team size. First, we predict that estimators will exhibit the *team scaling fallacy* -- they will not be sufficiently sensitive to the effect of increasing team size on the total amount of labor required to complete a project.

Hypothesis 1: The actual effort required to complete a project will exceed the estimated effort by a greater amount for larger teams than for smaller teams.

This hypothesis is based on the notion that people fail to fully appreciate the process losses associated with increasing team size. This is not merely a self-serving bias and thus should be exhibited by both the members of teams completing the projects (internal estimators) and by outsiders to those teams (external estimators).

We also predict that this bias is at least partially driven by a tendency to pay more attention to efficiency gains relative to efficiency losses.

Hypothesis 2: The more estimators focus on process gains relative to process losses, the more pronounced will be the team scaling fallacy that they exhibit.

The remainder of this paper is devoted to empirical tests of these hypotheses. First, we report on a laboratory experiment that documents the team scaling fallacy. Second, we replicate this finding in another laboratory experiment using new stimuli and show that it is associated with the tendency for estimators to focus more on process gains than process losses as teams expand. Finally, we document the team scaling fallacy in a field study of project forecasts at a large, international software company.

Experiment 1: Do Forecasters Exhibit the Team Scaling Fallacy?

In our first experiment, teams of varying size completed a project that was divisible among team members but required the integration of all the work into a single, final product. Estimators, who were not part of any team, were familiar with the kinds of workers who would staff the teams, but did not know any specific team members. This task was designed to model the way in which team projects are structured and forecasts are generated in many major industries (e.g., software design, consulting, and construction).

Team Exercise Participants

Two hundred sixty-seven executive MBA students (97 female, 170 male) at UCLA took part in a team exercise during a required class in organizational behavior. The exercise was completed for a class in which students were accustomed to completing activities whose purposes were explained only during end-of-class discussions. Thus, no cover story was given to the students about the exercise, and no material relevant to the team scaling fallacy was covered in class prior to the exercise. Participants were randomly assigned to either a two-person team

(34 teams) or to a four-person team (33 teams). All of the teams were then asked to complete a group construction project.

Team Exercise Procedure

In selecting a team exercise, we sought a task that was easy to learn (so it could be completed during a single class period), easy for estimators to conceptualize, involved opportunities for division of labor, and required coordination among team members to integrate their work. The task of assembling LEGO blocks into a pre-set structure met all these criteria, and is a team project that has been used by other researchers (Bluhm, Widiger, & Miele, 1990; Woolley, 1998; Zaccaro, Foti, & Kenny, 1991). All of the teams were required to assemble 50 LEGO pieces into a human figure ("LEGO Person").

Prior to the start of class, the instructor briefly presented an image of the LEGO Person on the classroom screen, then placed a single model of the LEGO Person at the front of the room. Students were told: "Your task is to assemble these pieces exactly like the model in the front of the room." Students were then given the names of their teammates (randomly assigned), as well as exercise instructions and LEGO sets. Finally, students filled out a short survey about their demographic characteristics and past experiences with members of their team.

Teams were then given up to 30 minutes to plan how they would assemble their LEGO pieces. Only during the planning period could group members observe the LEGO Person model, and even then, it could only be observed by one person at a time. Once a team decided it was ready to begin assembly (or once the 30 minutes available for planning had expired), the team began a timed assembly process. Whenever a team finished construction, its work quality was checked against the desired output. If the work did not precisely match the LEGO Person model, then one member of the team was given an opportunity to view the model again and report back to the team. The team was then given an opportunity to correct the defects, all with a timer running. Once the team's output matched the LEGO model, the team's total minutes of labor were recorded. Due to class length constraints, the assembly exercise ended after 30 minutes. Any teams that had not finished the project by that time were automatically assigned 30 minutes as their assembly time (twenty teams did not finish assembling the LEGO model).[2] Participants were debriefed on the purpose of the exercise and taught about the concept of coordination neglect after the exercise was completed.

Estimation Exercise Participants

One hundred seventy-eight undergraduate students were recruited through campus advertisements at the University of Pennsylvania to complete an estimation exercise. They were paid $10 apiece to complete a series of surveys over the course of an hour. Students received additional bonus pay up to $2, based on their performance on the estimation exercise (the bonus algorithm is described below). Upon beginning our study, which was described as a "decision making study", participants were told that "the purpose of the study is to learn about how people estimate unknown quantities."

Estimation Exercise Procedure

The estimators worked entirely independently (indeed, on a different university campus) from the teams assembling the LEGO pieces. However, estimators were shown the same instruction sheet and (picture of) the assembled LEGO Person model that was seen by participants involved in the construction task. Estimators were informed that executive MBA students at UCLA would be completing this LEGO construction project. They were then asked to estimate the total minutes of labor required to complete the LEGO construction project by a team composed of *either* two students *or* four students. Estimators were randomly assigned to one of the two team-size conditions and asked the following question:

> "Assume that teams 'billed' for their time (per minute) spent on the project per group member. So, a group of 3 that spent 10 minutes on the project billed for 30 person-minutes. Please forecast the average total number of person-minutes that it took a group of 2/[4] people in UCLA's executive MBA program to complete the assembly of the Lego Person."

Estimators were informed that they would be paid for accuracy. Specifically, they were given the following description of their incentive pay:

> "If your estimate of the average number of person-minutes billed by a 2/[4] person team at UCLA is exactly right, you will receive \$2.00 in bonus pay. For every minute separating your estimate from the correct number, your \$2.00 bonus will be reduced by \$0.05. So, for example, if your estimate is off by 10 minutes, your bonus pay will be \$1.50."

Results and Discussion

Summary statistics describing the actual and estimated total minutes of labor needed to complete the projects are provided in Table 1. The first column shows that four-person teams required more person-minutes to complete the project than did two-person teams ($t(65) = 2.17$, $p < .05$, one-tailed). We also examined how team composition affected project completion times. We did not see a statistically significant relationship between the average age of team members or team familiarity and construction speed. We do see that homogenous teams in terms of gender (i.e., all male or all female) complete the project faster than heterogeneous teams ($t(65) = 1.76$, $p < .05$, one-tailed).

The second column shows that estimates of the required person-minutes for project completion were also larger for teams of size four than for teams of size two ($t(176) = 3.00$, $p < .01$, one-tailed). Apparently, people are aware that larger groups require more total effort to complete projects.

To examine whether participants exhibited the team scaling fallacy, we created a variable designed to capture each estimator's degree of over-optimism. Specifically, we calculated the average of the *actual* person-minutes required by UCLA students to complete the project, in a team of two or four people, minus a University of Pennsylvania participant's *estimate* of the person-minutes required by such a team (Column 3 of Table 1). The larger this quantity, the more over-optimism an estimator had exhibited. We found that optimism was significantly greater for larger (4-person) teams than for smaller (2-person) teams ($t(176) = 1.76$, $p < .05$, one-tailed), consistent with Hypothesis 1. Thus, although estimators recognized that larger teams would require more time to complete the LEGO assembly project, they were still relatively insensitive to the impact that team size can have on the total amount of effort that a project requires.

Experiment 1 confirmed our first hypothesis in a controlled laboratory environment with random assignment of team size where external estimators were paid for accurate performance using a between-subject estimation design. We now turn to a replication using different stimuli in a within-subject estimation design involving a procedure that allowed us to test our second hypothesis.

Experiment 2: Does Focus on Process Gains vs. Losses Affect the Team Scaling Fallacy?

In Experiment 2, we attempted to replicate the team scaling fallacy found in Experiment 1 using a similar construction project, but this time eliciting estimators' views about the relative importance of process gains versus losses as team size increases. Experiment 2 thus allowed us to test our second hypothesis concerning a possible mechanism driving the team scaling fallacy.

This new experiment relied on within-subject estimation (participants estimated the number of person-minutes required to complete a project by teams of size two *and* size four, rather than by teams of one size *or* the other), to confirm that our findings hold in both joint and separate estimations. Although one could argue that estimates of the total amount of labor required by teams to complete projects are often generated one-at-a-time, there is a benefit to examining whether within-subject estimates exhibit the same patterns as between-subject estimates. Estimators in some firms make staffing decisions or are called on to budget work simultaneously by several teams of varying sizes. Moreover, a within-subject paradigm allows for a more conservative test of the team scaling fallacy because past studies have found that attributes difficult to evaluate (in this case, the impact of team size on effort required) receive greater weight when options are evaluated simultaneously than when options are evaluated separately (Hsee, Loewenstein, Blount, & Bazerman, 1999). Joint evaluations tend to be more thoughtful and less instinctive than separate evaluations, leading to less biased estimates (see Milkman, Rogers, & Bazerman, 2008, for a review). Because the outcomes of joint and separate evaluations have been shown to differ in past research, and because many estimates in the field can be made either jointly or separately, it is important to understand whether the team scaling fallacy occurs in both joint and separate evaluation.

Finally, in this experiment we elicited estimates of the total amount of labor required to complete projects not only from outsiders, but also from members of actual project teams. This allowed us to see whether people completing a project themselves are also susceptible to the team scaling fallacy. Comparing results from these two groups of estimators revealed the impact that personal stakes (and perhaps greater familiarity with the people completing the task) might have on the team scaling fallacy.

Team Exercise Participants

Eighty students (29 female, 51 male) at the University of North Carolina at Chapel Hill (UNC) took part in a team exercise during an MBA elective class. Because the exercise was for a class and students were accustomed to taking part in class exercises where the explanation followed the activity, no cover story was given to the students about the exercise, and no material relevant to the team scaling fallacy was covered in class prior to the exercise. Participants were randomly assigned to either a two-person team (12 teams) or a four-person team (14 teams). All of these teams were asked to complete a group construction project.

Team Exercise Procedure

All of the teams assembled a 188-piece LEGO set (Smash ‘n’ Grab, LEGO Item #5982). At the start, students were provided with an instruction sheet that included a picture of the fully assembled LEGO set. They were then told that “the objective is to assemble the LEGO pieces as quickly and accurately as possible.” Before beginning their work, students completed a short survey in which they estimated the total effort that would be required for teams of two people and four people (respectively) to complete the project. The order in which students answered questions about these two team sizes was counterbalanced. The students responded to the same question posed to estimators in Experiment 1. After completing this survey, students were given the names of their teammates (teams were again composed randomly) and permitted to meet with them. Students thus were unaware of their team’s composition or size when completing the estimation task.

Next, all teams were provided with a LEGO set and asked to begin their project. Their work was timed, and when a team finished its work, the result was checked against the desired output. If a team's LEGO structure did not precisely match the picture on the LEGO box, then the team was given an opportunity to correct the defects, with a timer running. When the team's output matched the LEGO model, the team's total minutes of labor were recorded. Unlike the first exercise, sufficient time was provided for all groups to successfully complete the exercise. After the exercise, participants were debriefed. We found that students were surprised to learn about the phenomenon of coordination neglect and eager to learn more about it. They did not appear to realize that we had been exploring this phenomenon in the construction exercise they had just completed. During the debriefing, students engaged in an animated discussion of their own experiences working in large, inefficient teams.

Estimation Participants

One hundred ninety-seven undergraduate students were recruited through campus advertisements at the University of Pennsylvania. These students were paid $10 each to complete a series of surveys over the course of an hour. Again, our survey was described as a "decision making study", and participants were told that "the purpose of the study is to learn about how people estimate unknown quantities."

Estimation Exercise Procedure

The external estimators worked independently (on a different university campus) from the teams of UNC MBAs assembling the LEGO pieces. Estimators were shown the same instruction sheet and picture of the assembled LEGO set that were shown to the students who were actually involved in assembling the Lego pieces. The estimators were told that the assemblers were MBA students at UNC. Then, the estimators were asked the following question about a team of either two or four persons:

"Assume that teams will bill for their time per minute spent on the project per

> group member. So, a group of 3 that spent 10 minutes on the project would bill for 30 person-minutes. Please forecast the average total number of person-minutes that it will take a group of 2/[4] people to complete the assembly of the Lego structure."

After responding to this question, estimators were asked the same question again, but for a group of whatever size they had not yet considered. The order of questioning was counterbalanced for the two group sizes.[3] Finally, participants were asked two questions about the rationale for their estimates. First, they were asked the following open-ended question:

> "If there was a difference in the assembly time estimates you provided for teams with different numbers of members, please describe your reasoning for estimating differences in assembly time as a function of team size:"

Next (to ensure that their answer to this open-ended question was unbiased), students were asked two questions about process intuitions. The order of these questions was also counterbalanced. In one question, students were asked to use a one-to-seven Likert scale to evaluate how the presence of additional team members among whom tasks could be subdivided affected their estimate. In the other question, students again used a one-to-seven Likert scale to evaluate how the presence of additional team members whose tasks must be coordinated affected their estimate.

Results and Discussion

Statistics summarizing the actual and estimated total minutes of labor required to complete the projects are displayed in Table 2. Column 1 shows that four-person teams took longer to complete the project than did two-person teams ($t(24) = 4.01$, $p < .01$, one-tailed).

Accuracy of external estimators. Surprisingly, for external estimators, estimates of the person-minutes required to complete the project for teams of size four did not differ significantly from those for teams of size two ($t(196) = 0.97$, NS, one-tailed). With respect to Hypothesis 1, the degree of estimator over-optimism (measured by subtracting estimated effort from actual effort for each project) was again significantly greater for larger (4-person) teams than for smaller (2-person) teams ($t(196) = 13.96$, $p < .01$, one-tailed). This result confirms that the team scaling fallacy plagues estimators even in a joint evaluation context.

Thus far, we have identified the team scaling fallacy by examining the difference between the *average* actual effort expended for a given team size and the estimated effort for that team size. However, using within-subject data, it is also possible to test for this bias by examining the accuracy of the ratio of an estimator's 4-person team effort estimate to her 2-person team effort estimate. We refer to this metric as the "4-to-2 effort estimate ratio". The benefit of calculating this ratio is that it provides a measure of each individual's sensitivity to team size - we can thus use it to explore what variables predict any given estimator's degree of bias (a question we will turn to shortly).

The actual number of person-minutes required for a four-person team to complete the LEGO project (111.50) divided by the person-minutes required for a two-person team (72.39) is 1.54. That is, four-person teams required 54% more person-minutes than did two-person teams to complete the project. We computed this ratio for each external estimator using the same approach, and then compared the estimated ratios to the actual ratio of 1.54. If estimators were not sufficiently sensitive to the effect of team size on the person-minutes required to complete a project, then we would expect their estimated ratios to be less than 1.54. Indeed, 84% of the external estimators had ratios below 1.54 ($p < .01$, sign-rank test, one-tailed), providing additional support for the existence of the team scaling fallacy.

Accuracy of internal estimators. We next turned to the effort estimates made by participants who actually completed the LEGO construction project (internal estimators). For this group, we found that estimates of the person-minutes required to complete the project were significantly greater for four-person teams than for two-person teams ($t(79) = 3.49$, $p < .01$, one-tailed).

To determine whether internal estimators also showed signs of the team scaling fallacy, we again created a measure of over-optimism by examining the *actual* person-minutes of effort in a condition (team size of two or four) minus an individual's *estimated* effort for teams in that condition. Once again, errors in effort estimates were significantly greater for larger (4-person) teams than for smaller (2-person) ones ($t(79) = 12.06$, $p < .01$, one-tailed). This result demonstrates that Hypothesis 1 holds not only when individuals make forecasts about tasks that others will engage in, but also when they make forecasts about tasks that they will engage in themselves.

Further evidence of the team scaling fallacy in internal estimations was found by comparing effort estimate ratios for 4-person teams with those for 2-person teams. The results showed that 75% of the internal estimators provided estimate ratios below the true ratio of 1.54 ($p < .01$, sign-rank test, one-tailed).

These findings demonstrate that the team scaling fallacy plagues both internal and external estimators, unlike some previously documented classes of biases.

Attention to process gains and losses as team size changes. Finally, we examined how external estimators evaluated the relative importance of process gains and losses when generating their effort estimates for teams of varying sizes. We did this in two ways.

First, we examined responses by the estimators to our open-ended question about the rationale for their estimates. Two research assistants who were blind to our hypotheses, and to the experimental conditions from which responses were drawn, were trained to code the estimators' responses. The coders were instructed to evaluate whether each response described (a) any benefits (in terms of efficiency, speed, etc.) that would be produced by increasing the size of a team, and (b) any costs (in terms of efficiency, speed, etc.) that would be produced by increasing the size of a team. The agreement rate for coding the first question was 80.2% (kappa = 0.57, $p < .01$) and the agreement rate for coding the second question was 79.7% (kappa = 0.57, $p < .01$). The coders discussed and resolved together any disagreements in their coding. The coded data showed that 107 participants (54.3%) mentioned coordination gains in their open-ended responses, whereas only 74 participants (37.6%) mentioned coordination losses. A two-sample proportion test showed that this difference was significant ($z = 3.34$, $p < .01$).

Next, we examined the external estimators' ratings of how important coordination gains and losses were for their estimates. The results showed that estimators considered coordination gains to be slightly more important than coordination losses ($\mu_{gain} = 5.45$ vs. $\mu_{loss} = 5.25$), although this difference was only marginally significant ($t(196) = 1.45$, $p < .10$, one-tailed).

Hypothesis 2 predicted that when estimators focused more on process gains than on process losses associated with greater team size, they would exhibit a stronger team scaling fallacy. To test this hypothesis, we first measured the extent to which each estimator exhibited the fallacy, using the 4-to-2 effort estimate ratio. We then took the natural log of that ratio, prior to running any regression models, in order to normalize the variable.

Table 3 reports the results from the analysis of the relationship between estimate accuracy and our two measures of attention to process gains and losses, the first captured by coding open-ended survey responses (Column 1) and the second through direct Likert scale ratings (Column 2). In both regressions, a negative coefficient corresponds to a larger team scaling fallacy (greater bias). Consistent with Hypothesis 2, both regression analyses showed a significant negative relationship between bias and the consideration of process gains, and a marginally significant positive relationship between bias and the consideration of process losses.

Using new stimuli, Experiment 2 thus replicated the finding that estimators underappreciate the additional effort required to complete a team project when team size increases, providing additional evidence of the team scaling fallacy. Further, Experiment 2 showed that these results persist in a within-subject estimation context (joint evaluation), as well as when participants estimate the effort required to complete team projects on which they themselves will work (internal estimation). By comparing the effort estimate ratios for 4-person teams with those for 2-person teams we can examine the strength of bias for internal versus external estimators. We find that the sample of external estimators are more biased than the sample of internal estimators in our study (i.e., their ratio is lower, $p < .01$, one-tailed). However, it is important to note that the populations of internal and external estimators compared here differed in a number of important ways besides whether or not they expected to complete the LEGO exercise themselves (e.g., external estimators were primarily undergraduates while internal estimators were MBA students), so these comparisons should be interpreted as merely suggestive rather than conclusive. Finally, the survey data collected in Experiment 2 supported Hypothesis 2 -- the team scaling fallacy is positively related to a heightened focus on process gains versus process losses in larger teams.

Field Study: Evidence of the Team Scaling Fallacy

Our field study extended the results of Experiments 1 and 2 by investigating whether the team scaling fallacy affects estimators in a natural setting. In particular, we examined whether project managers at a large international company were also more likely to underestimate the effort required to complete projects by larger teams. By examining the team scaling fallacy in the field, we could see whether this phenomenon persists when forecasters have considerable experience and receive frequent feedback about their forecasts, and when forecasting errors can have substantial financial consequences for a firm, as well as professional consequences for the forecaster.

Data Set

Our data set included observations of three years of software development projects completed for clients by a large software services company, SoftCo (a pseudonym). SoftCo develops customized software for many global customers. The company provided data to us because of an interest in understanding what factors affect estimation accuracy. Although SoftCo also performs other types of projects, such as software maintenance and testing, we examined only software development projects because performance and control variables were available for such projects, allowing for cross-project comparisons. SoftCo has a structured set of procedures and information technology systems for collecting data on project characteristics, team member experience, and project performance. We combined this information to create our data set.

Our data set consisted of 1,137 projects. Nineteen of these were missing some information and thus were excluded from our analysis. Among the 1,118 projects for which all information was available, we conducted our analyses controlling for (1) the “kilolines of code” (KLOC) produced by the team, which was a measure of project complexity (MacCormack, Verganti, & Iansiti, 2001), and (2) customer effects. Our final sample consisted of 594 software development projects because we removed 469 projects that did not track KLOC and 55 projects from customers for whom only one project was executed (given that part of our data analysis relies on a fixed effects regression model and therefore requires at least two observations per customer, all customer accounts with only one project were dropped from the analysis). These 594 projects were undertaken by SoftCo for 93 different customers a between 2004 and 2006.

Estimation of the total hours of labor required to complete a project at SoftCo takes place before the project starts and is carried out by sales and pre-sales professionals (not by the project team). Estimators receive role-based training that covers topics such as the "work breakdown structure" (WBS) and three-point estimation. In WBS, the work in a project is separated into the smallest practical pieces (PMI, 2008). Three-point estimation involves identifying the minimum, expected, and maximum effort that a task may require to complete. When creating a final estimate of the labor hours that a project will require, a SoftCo estimator has dual objectives. On the one hand, lower estimates may increase the likelihood of securing a contract from the customer. On the other hand, with higher estimates, SoftCo is more likely to deliver a project on-time and on-budget, thus increasing customer satisfaction and the odds of repeat business. Because SoftCo operates in a global, competitive environment, the marketplace prevents the firm from either adding too much slack to estimates (which would cause SoftCo to lose business to other firms with lower estimates), or estimating too aggressively (which would harm SoftCo's reputation). Thus, accurate estimates are highly valued at SoftCo.

The estimates we studied were forecasts of the total number of employee-hours that SoftCo expected projects to require for completion. After an estimate was made by SoftCo and approved by the customer, an execution team was created and staffed. Thus, estimators did not know exactly how many engineers would eventually work on a project (during the course of a project, the estimate of required employee-hours might change). Estimates are changed primarily because the customer changes the scope of the project (e.g., adding or removing requirements). In order to change an estimate, SoftCo requires both customer and internal approvals to make sure that inappropriate gaming is not taking place. We relied on final, revised estimates in our analyses because these most accurately encompass the final objectives of projects.[4]

To test for the team scaling fallacy among SoftCo estimators, we examined estimate quality -- the difference between the number of total employee hours estimated for a project's completion and the actual number of hours required to complete the project. To quantify estimate quality, we created a continuous variable that we called "% effort optimism." This variable was calculated by subtracting the estimated effort required to complete a project from the actual effort put into that project and then dividing the result by the estimated effort. In this field study, we standardized our performance measure, because unlike the laboratory experiments, we were comparing projects that involve different types of work. For that reason, we also included a number of project characteristics as control variables in our regression analyses. Given our expectation that estimators at SoftCo would also exhibit the team scaling fallacy, our prediction was that team size at SoftCo (measured by a count of the total number of individuals ever involved in a project) would have a positive association with % effort optimism. Hence, holding all other factors constant, larger teams should be associated with a greater tendency to underestimate the effort actually required to complete a project.

We conducted regression analyses on the full project sample (N=1,118) and on a sample for which controls for KLOC and customer effects were available (N=594). For the second sample, we ran a Hausman test to determine whether to use a random-effects or a fixed-effects model. Because the Hausman test failed to reject the null hypothesis (that a random-effects model was consistent), we used a generalized least squares (GLS) random-effects regression model (Greene, 2003). However, all of the results reported here are robust to replacing our customer random-effects models with customer fixed-effects models. We ran all models with robust standard errors clustered by customer.

Our data set included a number of variables that described both project characteristics and team human capital characteristics, all of which were included as controls in our regression models. These control variables are described in Table 4, and Table 5 provides summary statistics for each variable in the analyses.

Across the full sample of projects studied, the median software development project at SoftCo involved a team of 12 people; the smallest project was executed by 2 people, the largest by 151 people. The median project took approximately six months to deliver, requiring 5,451 hours of employee effort to complete. Project duration ranged from 23 to 1,082 days; project effort ranged from 256 to 268,253 employee hours. The median % effort optimism *score* was -3.9%, and this variable ranged from -63.0% to 274.1%.[5]

Results

Table 6 presents the results of regression analyses examining the relationship between team size and % effort optimism for the full sample of projects (Column 1) and for the restricted sample of projects where additional control variables were available (Column 2). In both cases, the coefficient for team size was positive and significant, implying that larger teams were associated with greater effort underestimation. In fact, a one-standard-deviation increase in team size was related to a 3.5% increase in effort optimism.

One concern with this result is that our measure of team size—a count of the total number of individuals ever staffed on the team—is imperfect. Ideally, we would prefer data on the team size that was expected when SoftCo began the project, but this information was not recorded at

SoftCo. Projects at SoftCo follow various staffing patterns, depending on their objectives (e.g., staffing everyone at once vs. gradually adding individuals). Looking at the maximum number of individuals ever involved in a project controls for these differences. However, it could overstate a negative effect of team size on % effort optimism because adding people to a project in its latter stages may exacerbate an effort overrun due to the added coordination issues (Brooks, 1975). Therefore, we ran robustness checks in which we excluded people staffed during the later stages of a project from our team size predictor variable. To do this, we counted the total number of team members who worked on a project at the midpoint of its duration. We also counted the total number of team members on a project at the 80th percentile of the project's duration. We then repeated all analyses, substituting these two team-size variables for our primary predictor, and found the same fundamental pattern of results reported earlier. We also repeated all of our analyses using both a log transformation of team size and the square of team size, to detect any nonlinear effects. Both models again supported the results reported above.

Discussion

Our findings suggest that estimators are not sufficiently sensitive to the effect of team size on the hours of labor required to complete a software development project. These results are consistent with the findings from Experiments 1 and 2 and provide additional evidence of the team scaling fallacy. Specifically, we found that % effort optimism correlated positively with team size, even after controlling for estimated person-hours to complete a project. The R-squared values in our regression models were small. But as a SoftCo manager told us, "Even a couple percentage points is very important. That money goes straight to the bottom line and can be the difference in a profitable or an unprofitable project. Also, if you get the estimate right, then you don't have to stress the team out in hitting the schedule."

Our field data illustrate the importance of the team scaling fallacy in a natural context where opportunities abound for learning and feedback and the economic consequences for inaccurate forecasts are substantial. However, we were not able to observe planned team size—only actual team size. As a result, we cannot conclusively eliminate the possibility that some projects were behind schedule and added team members as a result — thus creating an operational problem that contributed to poor performance. Although our robustness checks examining alternative measures of team size helped to address this concern, and also supported our main findings, we acknowledge that there could be some underlying, unobserved heterogeneity in teams driving our results. However, such an explanation cannot account for the converging results from our two laboratory experiments, in which we were able to exogenously assign project-team size while holding all other characteristics of a group task constant. Future work might examine the team scaling fallacy in a field experiment where team size can be manipulated.

General Discussion and Conclusion

In three studies, we found evidence that the team scaling fallacy is a real and persistent bias. Optimistic errors in forecasting the total labor required to complete team projects increase as the size of the project team increases. This bias is exhibited by both outsiders estimating the effort required to complete a project and by insiders who will be completing the project themselves. The bias is evident whether estimates about teams of varying size are made between- or within-subjects. We also found that the bias was more pronounced when estimators placed more weight on process gains and less weight on process losses in teamwork. Our research thus provides the first direct empirical support for a specific implication of the coordination neglect hypothesis proposed by Heath and Staudenmayer (2000).

The studies in this paper have complementary strengths and weaknesses. Our experiments, carried out in a laboratory setting, allowed us to examine the causal impact of exogenously varied team size on estimation error, while holding each team's task constant. The experiments also allowed us to explicitly measure estimators' intuitions concerning process gains and losses, which allowed us to document a possible mechanism for the team scaling fallacy. However, the experiments did not allow us to establish the external validity of our findings. Meanwhile, our field study established the external validity of our findings, but did not allow us to conclusively establish a causal relationship between team size and optimistic estimation error, or to examine the mechanism responsible for our findings. Together, the three studies provided compelling evidence that the team scaling fallacy is a meaningful phenomenon associated with a tendency to underestimate process losses and/or overestimate process gains.

Questions for Future Research

Our work suggests several questions for future research. First, it remains to be seen how team member interdependence affects the team scaling fallacy. If a team project requires little coordination, then coordination neglect is unlikely to bias estimators. Coordination costs vary across team projects, depending on task and team member interdependence. In this paper, we examined estimates of the total hours of labor required to complete projects that required significant coordination among team members. Future research could examine the sensitivity of the team scaling fallacy to interdependence of partitioned task components. For instance, one intriguing avenue of inquiry would be to examine the impact of increasing project modularity on the team scaling fallacy. By dividing work into modules that interconnect only through well-defined interfaces, modularity decreases the global consequences of local changes and thus limits both the number of interconnections and the need for them among team members (Baldwin &

Clark, 2000). Thus, modularity may reduce potential process losses sufficiently to attenuate or eliminate the team scaling fallacy. In any case it would be interesting to examine the extent to which estimators are sensitive to task modularity when making their forecasts.

A second question raised by the present work is how the team scaling fallacy might be affected by one's familiarity with the team completing a project. In our research, estimators were generally unfamiliar with the specific individuals who would be working on team projects. It would be interesting to understand how estimates change if an estimator knows, for example, whether the people on a team have worked together previously, or the demographic diversity of team members. Such factors have been shown in previous studies to affect team performance (Espinosa, Slaughter, Kraut, & Herbsleb, 2007; Harrison & Klein, 2007; Huckman, Staats, & Upton, 2009), but it remains to be seen whether estimators are aware of these effects, and to what extent estimators incorporate them into effort forecasts.

Although we focused broadly in this paper on the impact of attending to process gains rather than losses associated with increasing team size, it would also be interesting to examine estimator sensitivity to specific kinds of process gains and process losses in group work (for example, motivation problems and conflicts among members). Future work could explore which specific factors contribute to the team scaling fallacy, and compare the relative contributions of each factor. Future research might also examine whether characteristics of the estimator or the estimator's setting moderate the team scaling fallacy. For example, although we provided an accuracy bonus to all of the estimators in Experiment 1, future research could examine whether stronger incentives increase accuracy. In field settings, authority or status sometimes depends on accurate estimation, so it could be interesting to explore the impact of *non*pecuniary incentives, such as social accountability, on the team scaling fallacy. Further, we examined the team scaling fallacy in the context of individuals' estimates of group performance. But groups of estimators also could be studied. Would groups exhibit the team scaling fallacy as well (cf. Buehler et al., 2005; Sanna et al., 2005)? Finally, future research could consider the team scaling fallacy in the context of not only effort estimation, but also schedule estimation (i.e., missed deadlines).

Future research might also investigate the team scaling fallacy in other field settings. Although the archival dataset we examined was both large and detailed, it came from a single company in a single industry. Prior studies of small groups have shown that shifts in team size have implications for team performance across a wide range of settings (Hackman & Katz, 2010; Moreland et al., 1996). Thus, we would expect the team scaling fallacy to arise across a wide range of settings as well.

We found that the team scaling fallacy is associated with greater attention to process gains relative to process losses. One additional avenue for future research is to explore the extent to which this observation is driven by overweighting of process gains relative to process losses, or overestimation of process gains relative to process losses. That is, to what extent do people incorrectly estimate the magnitude of process gains and losses associated with increasing team size, and to what extent do they underweight the impact of these factors? Our data did not permit us to disentangle these different possibilities; future research could follow up on this question.

Implications for Project Management

A natural question about the team scaling fallacy is how it can be remedied. It might be helpful to measure and model the relationship between increasing team size and the number of hours of labor required to complete team projects in particular settings. Such information would allow estimates to be scaled by an appropriate, pre-determined factor, based on the expected team size, providing a cognitive repair for the team scaling fallacy (Heath, Larrick, & Klayman, 1998). Indeed, forecasters might benefit by estimating completion times using base rates that incorporate the true effect of team size on task-completion times, based on similar past projects (Kahneman & Lovallo, 1993).

Our findings have a number of direct implications for the field of project management. Although estimators and project managers are aware of the need for work integration within teams, our research shows the time required for such integration may be increasingly underestimated as a team grows. This problem may be compounded by common tools used in estimation, such as the work breakdown structure (WBS), which helps estimators to unpack the elements of projects that must be completed, so they can more accurately estimate total time required to complete the project. Benefits of a WBS focus around decomposition because "when the work is decomposed to greater levels of detail, the ability to plan, manage, and control the work is enhanced (PMI, 2008: 120)." Indeed, research on the planning fallacy has found that unpacking complex individual tasks into subtasks tends to increase the accuracy of predictions (Kruger & Evans, 2004). However, the WBS focuses solely on subdividing work, so there is risk that an estimator using the WBS may be particularly likely to neglect the *costs* of integration and team interaction. This could exacerbate the team scaling fallacy.

Both anecdotal and survey data suggest that organizations often struggle to accurately estimate the effort that will be required to complete the projects that they undertake. More work than ever is now being delivered to customers in the form of projects completed by fluid teams. Accurately estimating the costs of executing this type of work is therefore of growing importance (Edmondson & Nembhard, 2009; Huckman & Staats, 2011). Inaccuracy can have significant external consequences because missing promised budget and deadline estimates weaken potentially profitable projects and may result in lost repeat business (Heskett et al., 1997; Wheelwright & Clark, 1992). Additionally, although project leaders may attempt to meet deadlines by adding team members late in a project, or by asking existing team members to work longer hours, both solutions can have substantial negative consequences. Increased staffing may inflate labor costs and can still result in missed deadlines due to the time required for integrating new team members (Brooks, 1975; Chen, 2005). Likewise, overtime work can be an expensive solution due to both the higher rate of overtime pay and the risk that when team members work longer hours, stress levels can increase, resulting in worse performance and an increased threat of voluntary departure (DeMarco, 2002; Humphrey, Nahrgang, & Morgeson, 2007). We hope that our research on the team scaling fallacy can lead to new procedures and tools that improve the accuracy of estimates of group effort and minimize negative organizational consequences of errors in these forecasts.

References

Albanese, R., & Fleet, D. D. v. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management Review,* 10(2), 244-255.

Allen, T. J. (1977). *Managing the flow of technology*. Cambridge, MA: MIT Press.

Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: Volume 1*. Cambridge, MA: MIT Press.

Bazerman, M. H., & Moore, D. (2009). *Judgment in managerial decision making* (7th ed.). Hoboken, NJ: John Wiley and Sons, Inc.

Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization Science,* 14(3), 312-330.

Bluhm, C., Widiger, T. A., & Miele, G. M. (1990). Interpersonal complementarity and individual differences. *Journal of Personality and Social Psychology,* 58(3), 464-471.

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality & Social Psychology,* 50(3), 543-549.

Brooks, F. (1975). *The mythical man-month: Essays on software engineering*. New York: Addison-Wesley.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology,* 67(3), 366-381.

Buehler, R., Griffin, D., & Ross, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment (pp. 250-270*). New York: Cambridge University Press.

Buehler, R., Messervey, D., & Griffin, D. (2005). Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation? *Organizational Behavior and Human Decision Processes,* 97(1), 47-63.

Chen, G. (2005). Newcomer adaptation in teams: Multilevel antecedents and outcomes. *Academy of Management Journal,* 48(1), 101-116.

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science,* 4(1), 5-6.

DeMarco, T. (2002). *Slack: Getting past burnout, busywork, and the myth of total efficiency*. New York: Broadway.

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.

Edmondson, A. C., & Nembhard, I. A. (2009). Product development and learning in project teams: The challenges are the benefits. *Journal of Product Innovation Management,* 26(2), 123-138.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science,* 18(4), 613-630.

Gilovich, T., Griffin, D. & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive prediction.* Cambridge, U.K.: Cambridge University Press.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston: HBS Press.

Hackman, J. R., & Katz, N. (2010). Group behavior and performance. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology (Vol. 5, pp. 1208-1251*). New York: Wiley.

Hackman, J. R., & Vidmar, N. (1970). Effects of size and task type on group performance and member reactions. *Sociometry,* 33(1), 37-54.

Haleblian, J., & Finkelstein, S. (1993). Top management team size, CEO Dominance, and firm performance: The moderating roles of environmental turbulence and discretion. *The Academy of Management Journal,* 36(4), 844-863.

Hare, A. P. (1952). A study of interaction and consensus in different sized groups. *American Sociological Review,* 17(3), 261-267.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review,* 32(4), 1199-1228.

Heath, C., Larrick, R. P., & Klayman, J. (1998). Cognitive repairs: How organizational practices can compensate for individual shortcomings. *Research in Organizational Behavior,* 20, 1-37.

Heath, C., & Staudenmayer, N. (2000). Coordination neglect: How lay theories of organizing complicate coordination in organizations. *Research in Organizational Behavior*, 22, 153-191.

Heskett, J. L., Sasser, W. E., & Schlesinger, L. A. (1997). *The service profit chain: How leading companies link profit and growth to loyalty, satisfaction, and value*. New York: Free Press.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin,* 125, 576-590.

Huckman, R. S., & Staats, B. R. (2011). Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management,* 13(3), 310-328.

Huckman, R. S., Staats, B. R., & Upton, D. M. (2009). Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science,* 55(1), 85-100.

Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature. *Journal of Applied Psychology,* 92(5), 1332-1356.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From Input-Process-Output models to IMOI models. *Annual Review of Psychology,* 56, 517-543.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology,* 10(4), 371-384.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science,* 39(1), 17-31.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science,* 12, 313-327.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology,* 65(4), 681-706.

Knight, G. (2011, February). *Project management career opportunities in the healthcare industry.* Paper presented at the Healthcare Information and Management Systems Society Annual Conference. Orlando, FL.

Kruger, J. & Evans, M. (2004). If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology, 40,* 586-598.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin,* 89(2), 308-324.

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

Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment; Managing differentiation and integration*. Boston: Harvard Business School Press.

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

MacCormack, A., Verganti, R., & Iansiti, M. (2001). Developing products on "Internet time": The anatomy of a flexible development process. *Management Science,* 47(1), 133-150.

Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing our inner angels and demons: What we have learned about want/should conflicts and how that knowledge can help us reduce short-sighted decision making. *Perspectives on Psychological Science,* 3(4), 324-338.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review,* 115(2), 502-517.

Moreland, R. L., Levine, J. M., & Wingert, M. L. (1996). Creating the ideal group: Composition effects at work. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Small group processes and interpersonal relations*. (Vol. 2, pp. 11-35). Mahwah, NJ: Erlbaum.

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes,* 82(1), 117-133.

Mullen, B., Symons, C., Hu, L., & Salas, E. (1989). Group size, leadership behavior, and subordinate satisfaction. *Journal of General Psychology,* 116(2), 155-169.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition & the power law of practice. In J. Anderson (Ed.), *Cognitive skills & their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. M. (2006). The illusion of group productivity: a reduction of failures explanation *European Journal of Social Psychology,* 36(1), 31-48.

O'Dell, J. (1968). Group size and emotional interaction. *Journal of Personality & Social Psychology,* 8(1), 75-78.

Paulus, P. B., & Yang, H.-C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes,* 82(1), 76-87.

PMI. (2008). *A guide to the project management body of knowledge* (Fourth Edition). Newton Square, PA: Project Management Institute.

Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science,* 12(4), 502-517.

Sanna, L. J., Parks, C. D., Chang, E. C., & Carter, S. E. (2005). The hourglass Is half full or half empty: Temporal framing and the group planning fallacy. *Group Dynamics: Theory, Research, and Practice,* 9(3), 173-188.

The Standish Group: Chaos summary 2009. (2009). West Yarmouth, MA: The Standish Group International, Inc.

Stasser, G., & Taylor, L. A. (1991). Speaking turns in face-to-face discussions. *Journal of Personality & Social Psychology,* 60(5), 675-684.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185(4157), 1124-1131.

Watson, T. (2008). *Major NASA projects over budget*, from http://www.usatoday.com/tech/science/space/2008-03-25-nasaprojects_N.htm

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In G. Mullen & G. Goethals (Eds.), *Theories of group behavior* (pp. 185-208).

Wheelwright, S. C., & Clark, K. B. (1992). *Revolutionizing product development*. New York: Free Press.

Wooldridge, J. M. (2003). *Introductory econometrics* (Vol. 2E). Mason, OH: Thomson South-Western.

Woolley, A. W. (1998). Effects of intervention content and timing on group task performance. *The Journal of Applied Behavioral Science,* 34(1), 30-46.

Zaccaro, S. J., Foti, R. J., & Kenny, D. A. (1991). Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situations. *Journal of Applied Psychology,* 76(2), 308-315.

Footnotes

We thank management at the site of our field study for their commitment to this work. Max Bazerman, Chip Heath, Francesca Gino, and Christian Terwiesch provided valuable feedback on earlier drafts of this paper. Associate Editor Richard Moreland and four anonymous reviewers offered careful guidance that helped to improve this paper. This material is based on work supported by the National Science Foundation under Grant No. 0943210. Any errors are our own.

[1] There are $\frac{Nx(N-1)}{2}$ possible linkages within a team, where *N* is the number of team members.

[2] As an additional check, we drop all teams that did not complete the construction project in the allotted time. We still find that teams of size four took longer to complete the LEGO model, as compared to teams of size two.

[3] Analyzing the data revealed no order effects. The team scaling fallacy was neither stronger nor weaker for participants who considered the larger group first.

[4] As a robustness check, we repeated all our analyses using the original estimates. All results for team size remained qualitatively unchanged with this alternative outcome variable.

[5] Though SoftCo predictions are, on average, overly pessimistic, and the company is among the best performing in its industry. Note that the team scaling fallacy describes a tendency toward effort overoptimism for larger teams, but does not predict that all forecasts will be overly optimistic.

Tables

Table 1. Results and summary statistics from first construction exercise and related estimation task.

	Actual Time (person min) (1)	Estimates (person min) (2)	Difference (person min) (3)
	36.33	22.93	-13.40
Two-person team	(19.39)	(11.03)	(22.30)
	34 teams	89 estimators	89 observations
	52.64	30.07	-22.58
Four-person team	(39.11)	(19.53)	(43.71)
	33 teams	89 estimators	89 observations

Means, standard deviations (in parentheses) and sample sizes.

Table 2. Results and summary statistics from second construction exercise and related estimation task.

		Internal Estimators (UNC)		External Estimators (Penn)	
	Actual Time (person min) (1)	Estimates (person min) (2)	Difference (person min) (3)	Estimates (person min) (4)	Difference (person min) (5)
Two person team	72.39	34.07	-38.32	50.03	-22.36
	(23.81)	(15.19)	(28.24)	(57.48)	(62.22)
		80	80	197	197
	12 teams	estimators	observations	estimators	observations
Four person team	111.50	42.83	-68.66	47.12	-64.38
	(25.54)	(29.61)	(39.11)	(70.29)	(74.79)
		80	80	197	197
	14 teams	estimators	observations	estimators	observations

Means, standard deviations (in parentheses) and sample sizes.

Table 3. Results from analysis of relationship between process gain/loss survey responses and estimate accuracy using coding of open-ended responses (Column 1) and Likert scale responses (Column 2).

	DV: Log (Ratio of 4-to-2-person estimates)	
	(1)	(2)
Process gain	-0.42***	-0.10***
	(0.08)	(0.03)
Process loss	0.13*	0.05**
	(0.08)	(0.02)
Constant	0.04	0.14
	(0.07)	(0.20)
Observations	197	197
Adjusted R^2	0.24	0.08
F-statistic	32.28***	9.39***

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

All models include heteroskedasticity robust standard errors.

Table 4. Control variables included in the regression models.

Control Variable	**Explanation**
Team role experience	We average of the number of years that each member has spent at SoftCo in her hierarchical role when project manager, middle manager, or project engineer. Robustness checks with firm-specific experience generate the same pattern of results for this study's hypotheses.
Contract type	We use an indicator variable that is set to one when the contract is fixed price (where SoftCo is paid a set amount of money and bears any risk of overage) and is set to zero when the contract is time and materials.
Softco percentage	To control for coordination complexity due to employees working at different locations we construct a variable that divides the number of project hours completed at SoftCo's facilities by the total project hours.
Log (manual kilolines of code) Log (estimated duration) Log (estimated effort)	To control for project scale, complexity, and coordination challenges (Boehm 1981) we use three variables: the log of the number of kilolines of new source code that are written for a project, the log of the estimated duration and the log of the estimated effort.
Software language	Indicator variables for six software languages that appear in the data.
# of software languages	Indicator variable that is equal to one when a project has more than one software language and is equal to zero otherwise.
# of technologies	Indicator variable that is equal to one when a project employs more than one technology and is equal to zero otherwise.
Start year	Indicator variables for the year when the project started.

Table 5. Summary statistics and correlation table describing dependent, independent, and control variables of interest (*N* = 1,118 except for Log (kilolines of code) where *N* = 594).

	Variable	Mean	σ	1	2	3	4	5	6	7	8
1.	% effort optimism	-4.41	20.12								
2.	Team size	18.40	18.47	.03							
3.	Log (estimated effort)	8.74	1.11	-.07	.71						
4.	Log (actual effort)	8.65	1.11	.07	.71	.97					
5.	Team role experience	1.25	.64	-.02	-.02	-.05	-.07				
6.	Contract type	0.66	.47	-.05	-.06	-.11	-.13	.15			
7.	Softco percentage	0.82	.18	.10	-.09	-.18	-.16	-.18	-.04		
8.	Log (kilolines of code)	3.48	1.28	.08	.55	.70	.71	-.11	-.07	-.10	
9.	Log (estimated duration)	5.30	.66	.04	.45	.74	.73	-.04	-.13	-.12	.49

Note: $r \geq .06$, $p < .05$.

Table 6. Results of regressions of *% effort optimism* on team size.

	Dependent Variable: % effort over-optimism	
	(1)	(2)
Team size	0.19***	0.12**
	(0.05)	(0.06)
Team role experience	-0.61	-1.77
	(0.98)	(1.71)
Contract type	-1.62	-1.68
	(1.56)	(1.40)
Softco percentage	6.56**	0.59
	(2.93)	(5.34)
Log (manual kilolines of code)		2.79***
		(0.81)
Log (estimated duration)	6.47***	5.58***
	(1.75)	(2.07)
Log (estimated effort)	-6.27***	-5.43***
	(1.49)	(1.58)
Constant	6.98	5.73
	(7.17)	(9.31)
Observations	1,118	594
Overall R^2	0.06	0.08

Notes: *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. Models include, but results are not shown for the following variables: number of languages, start year, software language, and number of technologies.