© *Academy of Management Discoveries*
2017, Vol. 3, No. 4, 382–403.
Online only
https://doi.org/10.5465/amd.2015.0097

TEAMS VS. CROWDS: A FIELD TEST OF THE RELATIVE CONTRIBUTION OF INCENTIVES, MEMBER ABILITY, AND EMERGENT COLLABORATION TO CROWD-BASED PROBLEM SOLVING PERFORMANCE

CHRISTOPH RIEDL[1]
Northeastern University and Harvard University

ANITA WILLIAMS WOOLLEY
Carnegie Mellon University

Organizations are increasingly turning to crowdsourcing to solve difficult problems. This is often driven by the desire to find the best subject matter experts, strongly incentivize them, and engage them with as little coordination cost as possible. A growing number of authors, however, are calling for increased collaboration in crowdsourcing settings, hoping to draw upon the advantages of teamwork observed in traditional settings. The question is how to effectively incorporate team-based collaboration in a setting that has traditionally been individual-based. We report on a large-field experiment of team collaboration on an online platform, in which incentives and team membership were randomly assigned, to evaluate the influence of exogenous inputs (member skills and incentives) and emergent collaboration processes on performance of crowd-based teams. Building on advances in machine learning and complex systems theory, we leverage new measurement techniques to examine the content and timing of team collaboration. We find that temporal "burstiness" of team activity and the diversity of information exchanged among team members are strong predictors of performance, even when inputs such as incentives and member skills are controlled. We discuss implications for research on crowdsourcing and team collaboration.

[1] Corresponding author.
We wish to sincerely thank Karim Lakhani and Kevin Boudreau for their input and collaboration at earlier stages of this work.

Editor's Comment

This well-written paper focuses on the phenomenon of crowdsourcing and asks the question: How might groups of individuals collaborate most effectively in a crowdsourcing setting to produce high quality solutions to problems? The paper describes a rigorous field study with random assignment of individuals to groups that seeks to examine the conditions that could facilitate a team's performance on a problem-solving task in a crowd-based setting. The "discovery" is that the temporal "burstiness" of the team members' contributions, which suggests some effort to coordinate attention to the problem, plays a highly significant role in influencing the quality of solutions that teams produce. As one of the reviewers noted "this paper is a perfect 'fit' for the *Academy of Management Discoveries.*" I wholeheartedly agree—it focuses on an important yet poorly understood phenomenon and reports on the results of a rigorous field study that provides potentially important insights into developing our understanding of that phenomenon. I highly recommend that all Academy members read this paper.

Frances J. Milliken, Action Editor

Within the past decade, there has been an explosion in the use of crowds for outsourcing innovation problems often organized as contests (Boudreau & Lakhani, 2013; Kittur et al., 2013). Although the use of contests to spur innovation is, in fact, a fairly old practice (Fullerton, Linster, McKee, & Slate, 1999), the wide availability of Internet technologies has greatly facilitated their rapid deployment, as well as access to a worldwide supply of contributors, as a means to broadcast problems and facilitate submissions (Afuah & Tucci, 2012).

Crowdsourcing, as it is frequently deployed, is driven by the assumption that problems can be decomposed into parts that can be addressed by widely distributed, independent workers and can benefit from a large sample of independently generated potential solutions (Howe, 2006; Jeppesen & Lakhani, 2010; Malone, Laubacher, & Dellarocas, 2010). Examples of popular online platforms for crowdsourcing include Innocentive (science and technology)–https://www.innocentive.com, Amazon Mechanical Turk (microtasks)–https://www.mturk.com, Kaggle (data science)–https://www.kaggle.com, TopCoder (software, algorithms, data science)–https://www.topcoder.com, Threadless (t-shirt designs)–http://threadless.com, and 99designs (logos, graphics)–https://99designs.com. When work is conducted with a low level of interdependence, such as in a crowdsourcing environment, then key drivers of solution quality will be the knowledge and skills of the people who develop the solutions and their individual motivation to exert high levels of effort to solve the problem at hand (Afuah & Tucci, 2012; Boudreau, Lacetera, & Lakhani, 2011; Malone et al., 2010). Accordingly, many online platforms are designed to accumulate the insights of skilled and highly motivated individuals in a manner that generally minimizes the need for collaboration and coordination among them. Indeed, some argue that, under a growing range of circumstances, crowds of independent thinkers and solvers can yield results that are superior to those of more traditional, interacting teams (Surowiecki, 2004). Consistent with these assumptions, recent work, especially in computer science, has focused on developing advanced tools for crowdsourcing that focuses on ways to break complex projects into parts and to facilitate their recombination in a manner that minimizes the level of required real-time interaction among collaborators (e.g., Kittur, Smus, Khamkar, & Kraut, 2011; Retelny et al., 2014).

Nevertheless, the results of crowd-based innovation are often inconsistent and disappointing, raising the question of whether results could be improved by introducing some level of collaboration (Majchrzak & Malhotra, 2013). This is aligned with the increasing evidence of the importance of collaboration to discovery and innovation in science and business (Cummings & Kiesler, 2007; Wuchty, Jones, & Uzzi, 2007), and the underperformance of co-acting groups that do not work with a sufficiently high level of emergent interdependence (Hackman, 2011).

So how might we use teams in a crowdsourcing setting to the best effect? The few studies that have begun to explore teams in a crowdsourcing context have typically done so in a way that eliminates or minimizes real-time interaction or coordination (e.g., Kittur et al., 2011; Retelny et al., 2014). In doing so, do they lose the potential benefits of high levels of emergent interdependence, such as the pooling of diverse inputs that provide the usual basis of creative team products (Caruso & Woolley, 2008; Wageman & Gordon, 2005)? Or would a high level of interdependence undermine the creativity that can come from independently generated solutions or the productivity gains from the efficiency of solo work? We report on a large-field experiment of team collaboration in an online crowdsourcing contest in which solvers are randomly grouped into "teams," that is, groups of workers expected to work interdependently and to share responsibility for collaborative output (Hackman, 1987) as well as share associated rewards. In this way, the teams were designed to be high in both task and reward interdependence (Wageman, 1995).

Two key questions that motivated our research are (1) does a high level of collaboration enhance the performance of teams of crowd workers, above and beyond the influence of individual skill and incentives? And, (2) would cash incentives (further) improve performance in crowd-based teams? Crowdsourcing is a context in which workers do not expect to work highly interdependently with others; we may see that teams that attempt to work more interdependently ultimately are not as efficient or creative as other teams that work in the typically more independent manner. Cash incentives may or may not help; traditionally, online contests have been able to attract better talent and greater effort when cash incentives are offered (Moldovanu & Sela, 2006). Incentives in a team setting, however, are notoriously difficult to design as individual-based incentives can undermine collaboration and team-based incentives can encourage free-riding and can be diluted by the need to be divided among members whose level of contribution may vary (Nalbantian & Schotter, 1997). Thus, in our study, teams are randomly formed and randomly assigned to an incentive condition (or a control condition without cash incentives), and we examine the degree to which the exogenous inputs of cash reward and member skill, along with the endogenous level of collaboration, drive performance.

If, in fact, individual skill and motivation created by cash incentives are the main drivers of performance, then emergent collaboration should not enhance, and may even detract from, the performance of these randomly composed, temporary, crowd-based teams. If, however, emergent collaboration is an important driver of performance, even when accounting for individual skills and incentives, this would suggest that the dominant approach to crowdsourcing for complex innovation problems could be improved by facilitating collaboration, which has important implications on how we design and manage crowdsourced problem solving moving forward.

We explore these questions in the context of a large field experiment conducted on a platform hosting online software development contests in which 260 participants were randomly assigned to 52 teams of five for a 10-day contest. Teams were randomly assigned to compete (or not) for cash incentives, and we examine the effects of individual skill, incentives, and emergent collaboration on team performance. Among the major contributions of this paper are insight into how to best leverage member skill and motivation in a crowdsourced problem-solving setting. In addition, we demonstrate new measurement approaches for gauging emergent collaboration via temporal burstiness of activity and information diversity. These measures could be of broad use to study a variety of technology-mediated collaboration settings for which observation of communication content and timing are readily observable. We provide access to source code in the R programming language via a GitHub repository so that others may easily reuse and build on these measures.[2]

BACKGROUND

Crowdsourcing

The term crowdsourcing, coined by Howe (2006) originally meant "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in an open call." Although variations on the concept have emerged and the term is used broadly, here we focus on the general process of issuing an open call to a group outside of traditional organizational boundaries and facilitating the contribution of submissions via an internet-based platform to address innovation and R&D problems. The "open call" (or "broadcast search" as it is often referred to in the economics literature) aspect of crowdsourcing is a key element that differentiates it from other models of open innovation (Arolas & Ladron-de-Guevara, 2012) in which members from an identified and bounded set of organizations work together (i.e., Du Chatenier, Verstegen, Biemans, Mulder, & Omta, 2009). The more fluid boundaries around participation in most applications of crowdsourcing introduce both the possibilities of more divergent solutions as a result of the greater diversity of the participants but also more difficulties in fostering collaboration among them for the same reason. A few variants on the crowdsourcing process have attempted to foster some level of collaboration among contributors, although in doing so these attempts tend to incorporate tools for automating hand offs and minimizing the need for real time coordination (e.g., Kittur et al., 2011; Retelny et al., 2014). For most of the use cases, crowdsourcing contributors work relatively independently (Kittur et al., 2013; Majchrzak & Malhotra, 2013).

Recent reviews indicate that the process used for eliciting and combining crowdsourced inputs is a relatively understudied area (Pedersen et al., 2013). There is a general sense that higher levels of collaboration would be helpful, at least for crowdsourcing solutions to more complex innovation problems (Majchrzak & Malhotra, 2013). However, one appeal of crowdsourcing for many contributors is the

[2] https://github.com/riedlc/AMD-Teams

availability of quick sources of income with relatively little coordination hassle (Kittur et al., 2013) which may lead teamwork to be unappealing in this context. Thus, an examination of whether the benefits of team collaboration observed in traditional organizational settings would generalize to a crowdsourcing environment will help shed light on these issues.

Team Collaboration

The ability of a team to perform has traditionally been thought to be driven largely by the abilities of the individual members, either in the form of general cognitive ability (Devine, 1999; Devine & Philips, 2001; LePine, 2003) or in task and organizationally relevant skills (Zenger & Lawrence, 1989). Thus, many of the models of team effectiveness have incorporated major components related to team composition (Gladstein, 1984; Hackman, 1987), and over the last few decades, an extensive focus on team diversity has developed (e.g., Chatman, Polzer, Barsade, & Neale, 1998; Harrison & Klein, 2007; Kilduff, Angelmar, & Mehra, 2000; Milliken & Martins, 1996; Williams & O'Reilly, 1998).

There is also an extensive research literature on collaboration in teams (Kozlowski, 2015), the importance of team processes (e.g., Ilgen, Hollenbeck, Johnson, & Jundt, 2005; Marks, Mathieu, & Zaccaro, 2001), interdependence (Wageman, 1995), and the potential benefits of collaboration for the quality of solutions produced (Hackman, 2011). Ongoing debate concerns the relative contribution of member ability and emergent collaboration to performance (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). In exploring emergent collaboration, two aspects of team process that have emerged as particularly salient in recent work relate to how members temporally coordinate or synchronize with one another (Ancona & Chong, 1999; Janicik & Bartel, 2003; Okhuysen & Waller, 2002; Zellmer-Bruhn, Waller, & Ancona, 2004) and the diversity of information team members share (Bell, 2007; Mello & Rentsch, 2015).

Interpersonal synchrony and burstiness. Research on behavioral mimicry and interpersonal synchrony demonstrates that we are much more likely to mimic the behaviors of those to whom we are close or aspire to be like and will achieve greater behavioral synchrony in relationships that are more cohesive (Cacioppo & Cacioppo, 2012; Wiltermuth & Heath, 2009). Individuals who are more attentive to social cues are more likely to achieve synchrony and cooperation with interaction partners (Krych-Appelbaum et al., 2007), as well as higher collective intelligence (Chikersal, Tomprou, Kim, Woolley, & Dabbish, 2017). But is synchrony important to globally distributed, short-term crowd-based teams? And how does it manifest?

It is likely that if synchrony does manifest, it is reflected in the flow of communication and work-products. One of the trickier aspects of coordinating groups online is managing the flow of communication. Human communication has been shown to have rich temporal structure (Barabasi, 2005). Although temporal patterns can be partially attributed to circadian and weekly rhythms (Malmgren, Stouffer, Motter, & Amaral, 2008), detailed analysis has shown that they have more fundamental origins (Karsai, Kaski, Barabási, & Kertész, 2012). In particular, human communication is known to be intrinsically bursty (Barabasi, 2005; Goh & Barabási, 2008) and to contain strong pairwise correlations of interaction times (Karsai et al., 2012). In other words, rather than a randomly distributed pattern of communications, there tend to be periods of high activity followed by periods of little to no activity.

The temporal patterning of activities is an important aspect of team effectiveness in any environment (McGrath, 1991). Synchronous interaction is an orderly process wherein verbal and nonverbal cues help regulate the flow of conversation, enable turn taking, provide feedback, and convey subtle meanings. In lean, asynchronous communication environments, the communication of nonverbal cues is hindered, feedback is delayed, and interruptions or long pauses in communication often occur (McGrath, 1991), which are difficult to interpret (Cramton, 2001). In an asynchronous discussion, typically many topics are active at the same time, with team members making contributions at different times, possibly on different topics. This pattern can increase information overload and may reduce the synergy of team members if there are no links among the responses. In addition, long time lapses between communication events can lead to discontinuous and seemingly disjointed discussions (Montoya-Weiss, Massey, & Song, 2001; Ocker, Hiltz, Turoff, & Fjermestad, 1995).

These observations suggest that a significant challenge faced by online teams is coordinating the

Author's voice:
Was there anything that surprised you about the findings? If so, what?

Author's voice:
How did the paper evolve and change as you worked on it?

temporal patterns of group behavior (McGrath, 1991; Warkentin, Sayeed, & Hightower, 1997), which has been shown to be critical to performance even in traditional teams (Gersick, 1989; Montoya-Weiss et al., 2001). Temporal patterns of coordinating are a particular challenge in crowd-based teams, as, in addition to being globally distributed, participants also tend to tuck their contributions in around the edges of their "regular" activities (Kittur et al., 2013), leading to an even less regular schedule than we would observe in global teams in work organizations.

We contend that groups will vary in the degree to which members attend to and align their activities with one another; in other words, some teams may be burstier than others. Consequently, we anticipate that the level of burstiness in a team will be indicative of the degree to which members are attending to one another's activities online as members who collaborate via online platforms can see the contributions and communications of others and make choices as to whether to respond in kind. In this way, greater burstiness will function similarly to higher levels of interpersonal synchrony or temporal coordination in face-to-face contexts and, thus, is likely to be associated with better performance.

Developing finer-grained measures of time-based behavior in teams would enable researchers to examine the degree to which crowd-based teams exemplify the same temporal dynamics frequently associated with performance in more traditional teams, such as developmental patterns characterized by a punctuated equilibrium (Gersick, 1989) or temporal entrainment to particular rhythms of operating (Ancona & Chong, 1999). We anticipate that those temporal dynamics will play less of a role in crowd-based teams because of the wide variety of time zones and daily life contexts that exert independent influence on individual team members; nevertheless, these temporal dynamics are possibilities to consider as we examine the development of bursty behavioral patterns.

Diversity of exchanged information. A second aspect of team process frequently tied to performance relates to the level of diversity of information exchanged within teams. One of the assumptions associated with the value of composing diverse teams is that doing so will lead to a greater diversity of information exchanged (Harrison, Price, & Bell, 1998; Mohammed & Ringseis, 2001; Phillips & Loyd, 2006). Organizational researchers have demonstrated that increasing teams' exposure to diverse information can enhance performance, especially on tasks that require creativity (Austin, 1997; Bantel & Jackson, 1989; McLeod, Lobel, & Cox, 1996). Tasks that require developing alternative solution approaches or creating plans of action are likely to benefit, especially from access to knowledge and abilities that are diverse, because that diversity can lead to a greater quantity of ideas (Chatman et al., 1998; McLeod et al., 1996; Milliken, Bartel, & Kurtzberg, 2003) as well as non-redundancy of ideas or perspectives in the group (De Dreu & West, 2001; Stasser, Stewart, & Wittenbaum, 1995).

Many online platforms are formed with the explicit goals of eliciting diverse pieces of information and providing a mechanism for integration and coordination to enhance group performance. Indeed, in some of the earliest innovations in Internet-based collaboration, the goals were to elicit a diverse array of contributions to a collection that could be made widely available (Malone et al., 2010). Some of the more recently developed collaboration platforms are intended for amassing a diverse array of independent or even mix-and-matchable contributions, such as mobile software applications on an apps market platform or product reviews on an opinions platform (Boudreau, 2012).

Although information diversity has been widely recognized as important to team performance (e.g., Cronin & Weingart, 2007; Dahlin, Weingart, & Hinds, 2005; Harrison et al., 1998; Lazer & Friedman, 2007; Phillips, Mannix, Neale, & Gruenfeld, 2004), compositional diversity in team membership is often treated as the proxy for informational diversity (Harrison et al., 1998; Mohammed & Ringseis, 2001; Phillips & Loyd, 2006). This is often due to the difficulty of measuring the diversity of information shared directly. While in lab-based studies, the manipulation and measurement of the diversity of actual information shared is possible, in field-based studies, the actual diversity of information communicated in teams is frequently a black box. Thus, methods for measuring actual information diversity via analysis of observed online communications would be a better means for examining such diversity's association with group performance.

Cash Incentives and Team Performance

Another key question for implementing teamwork in crowdsourcing settings is what role, if any, monetary incentives might play in stimulating higher quality work. Monetary incentives have long been the main mechanism for inducing high levels of effort in traditional organizational settings (Lazear, 2000; Prendergast, 1999). At times, they have been shown to increase the quantity, but not the quality, of work produced (Jenkins, Mitra, Gupta, & Shaw, 1998). Cash incentives also can crowd out intrinsic motivations (Deci & Ryan, 1985; Frey & Jegen, 2001), which are especially important in the case of creative problem-solving work (Amabile & Fisher, 2009; Stephan, 1996).

The use of monetary incentives in group work is fraught with difficulties as group-based incentives are often subject to free riding (Alchian & Demsetz, 1972; Lazear & Shaw, 2007). Creating reward interdependence in teams can enhance performance, but only if accompanied by highly cooperative work behavior, which reward interdependence alone does not guarantee (Wageman, 1995; Wageman & Baker, 1997). It also is important to avoid creating conflicts between individual goals and team goals; otherwise, the team goals are likely to be undermined. Consequently, the literature on the relationship between monetary (or monetary equivalent) incentives and group performance is mixed. Extrinsic, group-based rewards can enhance performance when they encourage more collaborative group behavior and when they serve to enhance individuals' intrinsic motivation. This "motivational synergy" is most likely to occur when people feel that the reward confirms their competence and the value of their work or enables them to do work that they were already interested in doing (Amabile, 1993). This is consistent with earlier research that demonstrates that "informational" and "enabling" rewards can have positive effects on intrinsic motivation (Deci & Ryan, 1985) and performance.

Incentives are widely used in crowdsourcing contexts. Although individuals participate in online contests for a variety of reasons (Boudreau et al., 2011; Malone et al., 2010), monetary incentives have generally been shown to enhance overall individual participation levels and quality of performance. In crowd-based teams, however, it is unclear whether monetary incentives will be a primary driver of performance, given the mixed effects on motivation described previously (Nalbantian & Schotter, 1997; Wageman & Baker, 1997). Thus, a field experiment that involves randomly assigned cash incentives and team member assignment can allow us to assess the causal effects of monetary incentives and team member skills in a setting that is not fraught with homophily and self-selection that often govern the formation of teams.

THE CURRENT STUDY

As described, the current literature on crowdsourcing, team-based cash incentives and team process, all yield inconsistent predictions about what may occur as a result of implementing real-time team collaboration with cash incentives in a crowdsourcing setting. Crowdsourcing and cash incentives are both more commonly implemented in the context of solo work by individuals, with a limited knowledge base (in the case of crowdsourcing) and mixed results (in the case of incentives) related to their implementation in team-based collaboration. Few settings permit a strong test of the relative contribution of team composition, incentives, and emergent collaboration processes to team performance. In this study, we conduct a field experiment on a crowdsourcing platform that regularly hosts contests for solving difficult algorithm and software development problems where the exogenous inputs of team composition and incentives can be randomized.

Research Design and Setting

Setting. Investigating the drivers of crowd-based collaboration is fraught with several empirical and methodological challenges. We chose to implement our research in a field setting that enabled us to obtain fine-grained measures of performance, skill, and behavior and to employ randomized assignment. The study was conducted over a 10-day online event held on a leading open software innovation platform. The online platform has a well-developed ratings and skills assessment system that enables the identification of the ability of any competitor based on the Elo rating system used in chess. Virtual workspaces were assigned to each team to enable messaging and communications and to enable software code development. During the time of the competition, we were then able to obtain detailed and objective measures of team member ability, communication, and team performance.

The algorithm problem that was being solved is not just representative of a typical (highly-challenging) algorithm development problem; it is, in fact, a real computational-engineering problem whose solution would be used by the aeronautics agency sponsoring the competition. The selected problem required the development of a robust software algorithm, which would recommend the ideal composition of the flight medical kit that is included on each flight. As mass and volume are severely restricted by the high costs of space flight, the medical kit has to be designed in a way such that as many expected and unexpected medical contingencies as possible can be met through the resources in the kit to avoid interrupting flights. The medical kit also has to be attuned to the characteristics of the flight mission and crew. Hence, the challenge was to develop a software algorithm, based on mission characteristics, which would minimize mass and volume of the medical kit and yet allow it to have resources such that the likelihood of a medical evacuation is minimized.

We collaborated with an aeronautics agency and the online platform in developing the problem statement and a test suite to generate an objective score for innovation problem-solving performance. The test suite included 100,000 flight scenarios, which were used to score solutions. In addition, the online platform worked with us to make various changes to their system and website so that we could run a controlled field experiment in their setting. For example, the platform was changed such that each team not only observed its own score but also could observe the scores of competing teams within their room on an ongoing basis (i.e., the highest score solution submitted up that point).

Incentive manipulation. Groups of four teams (20 individuals) were randomly assigned to virtual rooms; six of those rooms (24 teams) were randomly assigned a cash prize of $1,000 per room, awarded to the best of the four teams in each room. The remaining 28 teams did not compete for a crash prize within their rooms. All awards were *team-based* cash incentives as they would be shared across the five team members. For groups assigned to the cash prize condition, the division of the prize was based on an anonymous poll taken after the contest that asked each team member to allocate a percentage of the $1,000 to themselves and the other team members. Awards were based on the average allocation across all members of the team. Members of the highest scoring team in the entire exercise (i.e., across all rooms, including those not assigned a cash incentive) also were awarded the grand overall prize, which was a trip to see the launch of the flight vehicle in person (all five team members were invited to go). Furthermore, all contestants who actively participated received a contest t-shirt.

Participants. To recruit participants, an open call was issued to software programmers who had previously competed in contests sponsored by the platform, inviting them to enter a new contest sponsored by an aeronautics agency. In response, 260 elite algorithm developers entered and were randomly assigned to 52 teams of five. The participants were, to some extent, aware of the possibility of working in groups, but this was not the central message of the advertisement. We stressed that this would be a competition to solve a challenging algorithmic design problem. The usual mode of interaction on this platform was in the form of individuals' competing against one another to solve such problems. Once the contest began, they were informed that they would be working with a team. Inasmuch as individuals self-selected to join this contest, we should expect that they might have some inclination toward teamwork. In this sense, we might expect the participation levels that we observe in this context to, perhaps, be high in relation to what might be expected in a case in which an online platform attempts to encourage its members to coalesce into productive work groups under more typical operating conditions.

In the following sections, we detail the measures used in our empirical analysis and the methods used for analysis. In particular, we present the automated text analysis used to measure information diversity and the approach used to index the burstiness of team activity. Table 1 provides an overview of all of the variables used in our analysis.

Dependent variable: Team score. At the end of the competition, the last submission of each team (irrespective of which team member made the submission) underwent final system testing to determine an objective performance score of the algorithm, which is a measure of the quality of the algorithm that a team has developed. The score was computed as a function of the mass and volume of the medical kit (lower is better) and the number of evacuation scenarios that can be averted with the medical kit (higher is better).

TABLE 1
Variable Definitions

Variable	Definition
(1) Team score	The final system testing score of the last submission made by a team (i.e., an objective measure of team performance).
(2) Group cash incentive (exogenous)	Indicator whether a team was randomly assigned to cash treatment.
(3) Team skill (exogenous)	Mean skill rating (based on performance in prior contests) of randomly assigned team members.
(4) Information diversity (standardized)	Measure of diversity of information exchanged by a team.
(5) Burstiness (z-score)	Measure of temporal correlation of team activities.
(6) Collective team output	Number of all code submissions made by a team.
(7) Num communications	Number of messages exchanged within a team.
(8) Time zone index	Mean dyadic time zone difference of all team members.
(9) Num countries	Number of different countries represented in a team.

Exogenously varied: Group cash incentive. To index the random treatment assignment of the team-based cash incentive, we include a dummy variable, *Group Cash Incentive*, if a team was competing for the cash prize (coded 1) or was not competing for a cash prize (coded 0).

Exogenously varied: Team skill. To obtain an objective measure of the skill level of team members, we use a skill rating based on historical performance in prior competitions hosted on the same online platform. The individual-level skill ratings are based on the Elo chess-rating scheme and provide a reliable measure of skill and experience in algorithm contests. All participants in the competition had competed in at least one contest before participating in our experiment. We compute the team skill as the arithmetic mean of all five team members' ratings (alternative specifications that use maximum or median skill do not substantively alter our conclusions). Because individuals were randomly assigned to a team, the skill composition of a team was exogenously manipulated through the experiment.

Process variable: Information diversity. To index the effects of information diversity on team performance, we perform automated content analysis of all messages sent within each team. We perform the usual information retrieval preprocessing of eliminating stop words and stemming. Stop words are common words such as articles ("a," "an" and "the") and prepositions (e.g., "from," "of," or "to"). Stemming reduces words to their root; for example, "running" is recorded as its root, "run." That way, different inflected and derived words can be mapped to identical root words. Furthermore, we remove numbers from the text and restrict our text analysis to word stems of at least two characters in length. We then perform content analysis using latent Dirichlet allocation (LDA) to classify texts into distinct topics using a Gibbs sampler (Blei, Ng, & Jordan, 2003). LDA is an advanced statistical technique that is widely used in information retrieval and machine learning. It is a generative probabilistic model that allows modeling the topics of a corpus of documents such that each topic consists of a vector of words that are statistically related to each other.[3] The words related to a topic then can be ranked by their probability of appearing in a given topic.

LDA classifies topics in two distinct steps. In the first step, the entire corpus of all texts is used to discover the entire distribution of topics present across all teams. In the second step, each individual text is assigned probabilities for each topic. In our application, this provides us, for each message, with a distribution over topics that are most likely to have generated that message. We model 100 topics using the entire corpus of all 1,741 team messages (after preprocessing, 1,705 non-empty documents with 5,853 terms remain). Our analysis is robust to the number of topics used for the LDA topic model. We find almost identical results using a wide number of topics, ranging from 50 to 150.

Table 2 shows five sample topics from the resulting LDA topic model. The topics represent distinct areas of online collaboration, the problem domain of the contest, and the contest environment itself. For example, the topic "Medkit Contents" captures conversations that concerned the resources and their quantities that should be part of the medical supplies kit. Similarly, the topic "Sharing Code" contains terms related to the conversation about how team members could efficiently share programming code among themselves. The topic "Temporal Coordination" captures conversations related to when team members were planning to work on the algorithm contest and how they intended to coordinate their efforts over time. In summary, the topics capture relevant aspects of team communication that include different kinds of dependencies and challenges faced by the team and the coordination processes to overcome these challenges.

Once the topic model is constructed, we then compute an index of information diversity for each team as the average cosine dissimilarity of the topic space of each team's messages. This is a common measure that is widely adopted in information retrieval research to measure diversity of information. Specifically, we measure team i's information diversity ID through a normalized, squared sum of the cosine distance between the topic vectors of each of N messages d_j sent within the team and the mean topic vector M_i that represents all messages sent within team i:

$$ID_i = \frac{\sum_{j=1}^{N}\left(1-\cos\left(d_j, M_i\right)\right)^2}{N}. \quad (1)$$

The resulting measure is bounded by $0 \leq ID_\mathrm{i} \leq 1$. The measure aggregates the similarity of message vectors in a team from the mean topic vector in a team, thus approximating the spread or variance of

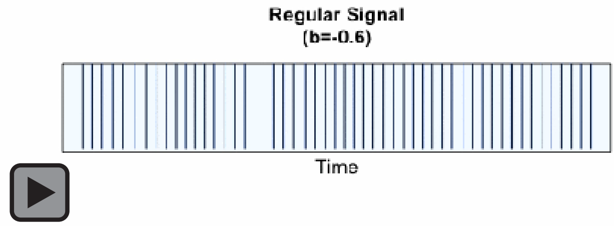
[3] Given a document corpus, LDA models each document d as a finite mixture over an underlying set of topics, where each topic t is characterized as a distribution over words. A posterior Dirichlet parameter $g(d; t)$ can be associated with the document d and the topic t to indicate the strength of t in d. For details of the algorithm, please refer to (Blei et al., 2003).

TABLE 2
Five Sample Topics from the 100-Topic LDA Model of Team Messages

"Medkit Contents"	"The Java Offline Tester"	"Sharing Code"	"Temporal Coordination"	"Contest Rules"
resource	java	attach	tomorrow	list
evacuation	tester	file	happen	contest
mission	offline	I've	hour	submission
amount	output	help	check	contribute
require	Xmx1024M	post	thought	end
medkit	run	edit	don't	prize
quantity	tester.jax	make	expect	allocation
number	program	got	time	group
maximum	source	use	please	memory
remove	jar	see	brief	since

topics in each individual message. Teams that send messages of diverse topics have higher information diversity scores than do teams that send messages of a more homogeneous topic distribution. One key feature of the measure is that it assesses information diversity relative to the topics covered within a team rather than for the entire topic space. Figure 1 schematically illustrates less and more diverse team communication. This measure of information diversity has been used in other studies that investigate communication behavior, most notably by Aral and Van Alstyne (2011). Our measure differs from theirs by applying more advanced LDA topic modeling techniques rather than support vector machines. The measure is included in regression models in standardized form.

Process variable: Burstiness of team activity. To index the temporal coordination of communication and code submissions within the team, we constructed a measure of the burstiness of team activity. This measure captures the degree to which team members concentrated their communication and work effort during relatively contained time periods versus spreading them out over time more equally. Specifically, we constructed a measure that captures the bursty nature of team activity based on the wait times (in minutes) between each team activity (either sending a message to the team or making a code submission). Greater correlation in the timings of team activities indicates greater burstiness. That is, greater burstiness indicates higher responsiveness of activity among members of the team. Conversely, if team activities are not well coordinated, this is equivalent to team activities' following a random Poisson process, resulting in a low degree of burstiness. Thus, low burstiness indicates that team activities are less temporally correlated. Figure 2 provides an intuition for the measure, using random and sampled data that illustrate cases of high and low temporal coordination.

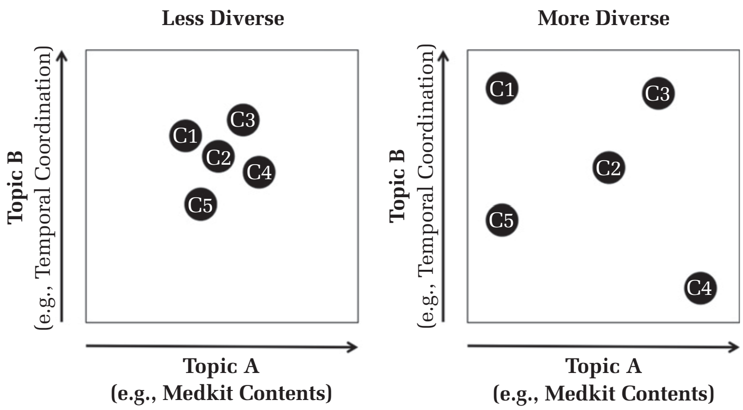


For each team i, we compute a coefficient of variation measure B (Goh & Barabási, 2008), defined as the ratio of the standard deviation to the mean of wait times τ between team activities

$$B \equiv \frac{(\sigma_\tau/m_\tau - 1)}{(\sigma_\tau/m_\tau + 1)} = \frac{(\sigma_\tau - m_\tau)}{(\sigma_\tau + m_\tau)}. \tag{2}$$

This measure is meaningful when both the mean and the standard deviation of wait times $P(\tau)$ exist, which is always the case for real-world finite signals (Wooten & Ulrich, 2017). B has a value in the bounded range $[-1, 1]$, and its magnitude indexes with the signal's burstiness: $B = 1$ corresponds to the most bursty signal, $B = 0$ to neutral, and $B = -1$ to a completely regular (periodic) signal. Thus, higher values of B correspond to spiked patterns of high team activity (high correlation of activity), whereas lower values of B correspond to more regular team activity (low correlation of activity). Because teams varied in overall activity (different total numbers of messages and code submissions per team), one concern might be that the measure is overly sensitive to the total number of team events. To address this concern, we compute bootstrapped null models of B scores using 1,000 random samples for different numbers of team events (i.e., the expected distribution of wait times and, hence, of B for a team that has N events). Based on these null models, we then compute a z-score, BZ, for each team. Figure 3 shows the empirical cumulative distribution of response times of team activities across all teams on a log–log scale. The points follow a straight line over several

FIGURE 1
Illustration of Less and More Diverse Team Communication



Note. Both teams draw equally heavy on Topic A and Topic B, resulting in similar average topic vectors. In the less diverse case (left), each communication draws to a similar degree on both topics. Communication C3, for example, contains about equal amounts of Topic A and Topic B, resulting in homogeneous topic usage. In the more diverse case (right), communications rely more heavily on specific topics. Communication C4, for example, consists almost exclusively of Topic A while communication C1 consists almost exclusively of Topic B, resulting in more diverse topic usage.

orders of magnitude, which indicates that the distribution of response times follows a power-law distribution. This supports the notion that team activity in our experiment follows the general busty patterns reported for other contexts of human communication (Barabási, 2010).

Control variable: Collective team output. We use the number of total code submissions made by a team as a control for team effort. Prior research has shown that, in tournaments that allow repeat entries, the number of submitted attempts is a strong predictor of performance (Wooten & Ulrich, 2017).

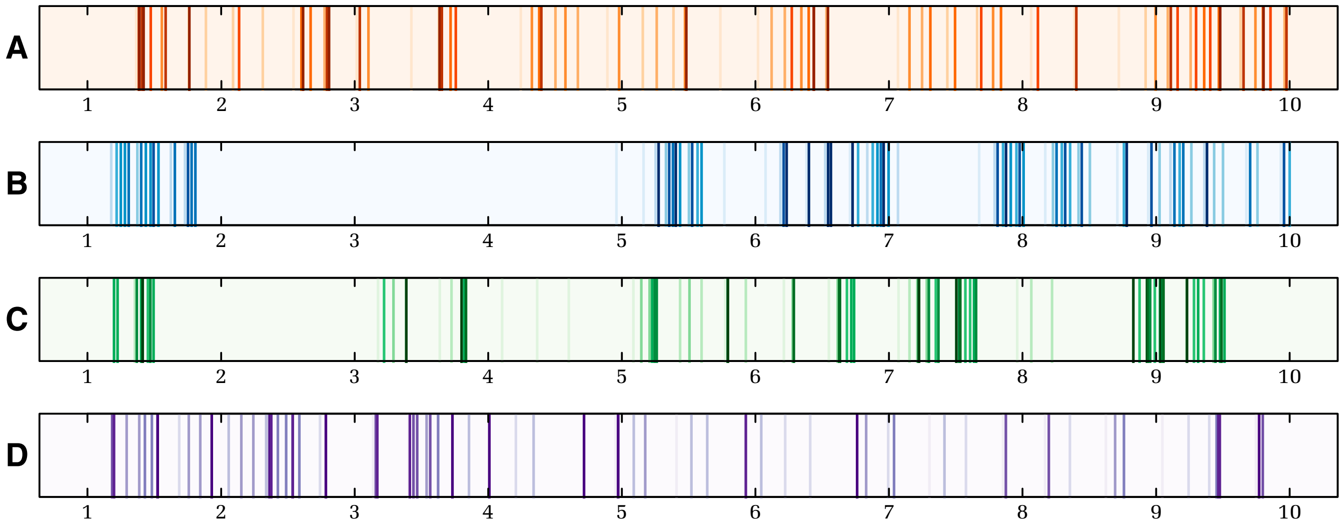
Control variable: Number of communications. The analyses of the content and timing of communication described previously were all conducted while controlling for the overall number of communications exchanged. Participants interacted on the platform through a web-based interface as is typical of such platforms. This screen included a workspace where they could read the problem statement, perform algorithm development (in the form of software coding; a number of software languages are possible), and submit solutions for compilation. Furthermore, individuals could communicate with other participants within their group via a bulletin board. That is, individuals could engage in general or directed communications with other group members by virtue of the content of their messages but communications were observable on a posting "board" or online forum visible to all group members (but not visible to other groups). Such a posting board was familiar to the platform as a general posting and discussion forum tool. The number of communications is indexed as the sum of all messages sent to the team message board.

Control variable: Time zone index. Research has shown that the dispersion of team members across time zones can affect the level of team coordination. For example, teams that span multiple time zones often experience significant challenges in coordinating schedules and deliverables (Massey, Montoya-Weiss, & Hung, 2003; O'Leary & Cummings, 2007; O'Leary & Mortensen, 2010). To control for the effect of temporal dispersion in our teams, we compute, for each team, a measure of pairwise time zone distances between all team members (i.e., we compute the *Time Zone Index* suggested by O'Leary & Cummings (2007)).

Control variable: Number of countries. Research has shown that compositional diversity with regard to team members' cultural backgrounds can affect team performance (Gibson & Gibbs, 2006; O'Leary & Mortensen, 2010; Reagans, Zuckerman, & McEvily, 2004). To control for the effect of cultural diversity in our teams, we control for the number of different countries represented in each team.

Sample demographics. We collected demographic information from our participants during registration and through a postexperiment survey immediately after the experiment was completed. The 260 participants came from 50 different countries, with over 50 percent of the competitors' coming from four countries: India (18 percent), China (13 percent), United

FIGURE 2
Illustrations of "Burstiness" and Difference between Activity Patterns



Note. Differences in activity patterns for teams with 90 team events with color-coded inter-communication times. Horizontal axis denotes time (day of competition), and each vertical line corresponds to a team event. The shorter the time between consecutive events, the darker the color. **a)** Sequence of regular, Poisson-distributed event times, which assumes that, at any given moment, an event takes place with probability q. Note that the inter-event times are comparable to each other and that long delays are virtually absent ($B = 0.00$). **b)** Sequence of events for a heavy-tailed (power-law) distribution. Note the very long delay times, followed by bursts of activity ($B = 0.54$). **c)** Sequence of typical activity pattern for a well-coordinated team, showing some longer wait times followed by bursts of activity, especially in the beginning, middle, and toward the end of the competition ($B = 0.36$). **d)** Sequence of typical communication pattern for a poorly coordinated team with an almost complete absence of bursts but, instead, regular event activity ($B = 0.00$)

States (12 percent), and Russia (9 percent). Figure 4 shows the geographic distribution of participants.

Table 3 shows the descriptive statistics for the main variables that we investigate. Participants were almost exclusively male (of the 122 participants who reported their gender, only five indicated female). Two-thirds of participants were students (165); the others were professionals (93; two declined to answer). Among those who reported being employed, the majority reported working in the software industry (63). The other industries represented by at least two participants were education/services (5), government/military (4), engineering (3), financial services (3), consulting (2), and R&D (2). Overall, the participant pool was young, with 161 who reported their age bracket as 18–24 years; 72, as 25–34 years; and 21, as 35 years or older (six declined to answer). Participants reported the following levels of education: secondary education (14), college or university (59), master's degree (38), and doctorate (11); a total of 138 chose not to report their level of education.

Empirical approach. In our empirical analysis, we consider the team as the level of analysis, and we have 52 independent data points, each corresponding to a single team. Our primary interest is to understand how exogenous variation in team-based cash incentives and team skill, as well as the emergent (endogenous) variation in information diversity and burstiness of team activities, affect team performance. Given that our dependent variable, team performance, is a continuous measure censored at zero, we employ Tobit regression (Kleiber & Zeileis, 2008). Statistical analysis was carried out in R (R Core Team, 2016). Code to compute information diversity and burstiness from communication data is available on GitHub.[4]

We estimate versions of the following model:

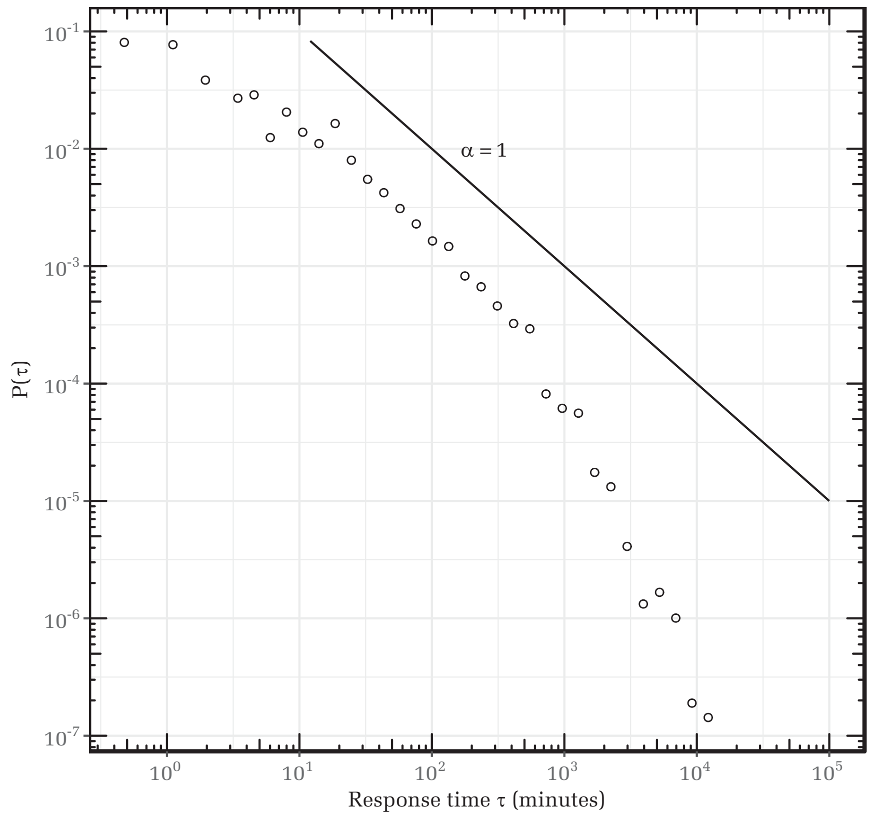
$$y_i = \alpha + \beta_1 \, Team\ Skill_i + \beta_2 \, GroupCashIncentive_i + \beta_3 ID_i + \beta_4 BZ_i + \beta_{5-9} \, Controls_i + \varepsilon_i. \quad (2)$$

These models estimate team i's performance y_i as a function of team i's exogenously determined team skill ($TeamSkill_i$), the randomly assigned team-based cash incentive ($GroupCashIncentive_i$), the team messages' information diversity (ID_i), the temporal correlation of team activity (BZ_i), a vector of controls, and an error term (ε_i).

Several empirical and methodological challenges arise when studying emergent collaboration. The

[4] https://github.com/riedlc/AMD-Teams

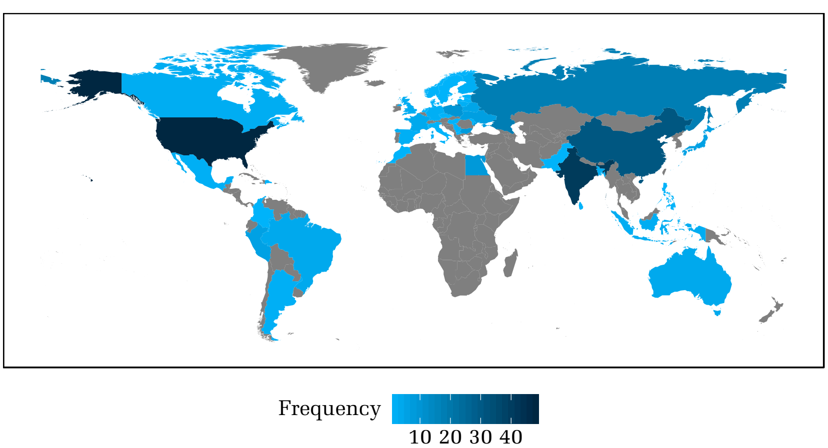
FIGURE 3
Bursty Nature of Team Activity



Note. If points plotted on a log-log scale follow a straight line over several orders of magnitude, this indicates a power-law distribution.

results that we present in the following section contain evidence to support several causal and noncausal inferences. The experimental treatment of team-based cash incentives is randomly assigned and, therefore, can be interpreted causally. For example, if performance in teams with team-based cash incentives is higher, we can conclude that this is a direct causal effect of providing monetary incentives to the team. The results for team skill also can be interpreted causally as team

FIGURE 4
Geographic Distribution of Participants



Note. Participants from 50 countries participated.

TABLE 3
Descriptive Statistics for Main Variables

	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Team score (1)	4170.53	3351.06	0.00	8952.79								
Group cash incentive (2)	0.46	0.50	0.00	1.00	0.27							
Information diversity (3)	0.04	0.05	0.00	0.20	0.67***	0.10						
Burstiness (z-score) (4)	2.07	2.16	−1.89	8.25	0.71***	0.18	0.78***					
Team skill (5)	−0.04	0.36	−0.78	0.66	0.42**	0.24	0.33*	0.45***				
Collective team output (6)	12.02	15.05	0.00	76.00	0.73***	0.34*	0.67***	0.71***	0.42**			
Num communications (7)	15.44	32.60	0.00	209.00	0.46***	0.03	0.60***	0.74***	0.40**	0.56***		
Time zone index (8)	4.77	1.76	0.40	7.20	0.12	0.01	0.18	0.14	−0.02	0.21	0.23	
Num countries (9)	4.19	0.84	2.00	5.00	−0.05	0.02	−0.08	−0.23	0.03	−0.04	−0.15	0.25

composition is random and, thus not biased by homophily or self-selection. The results from the analysis of emergent team processes, even in the presence of exogenously imposed team inputs are by their very nature, endogenous and should be interpreted as correlations, rather than as evidence of causality (at least without making additional assumptions).

The data from our study have three key features that greatly alleviate concerns common in other studies of team performance. First, investigating the drivers of collaboration and teamwork is often complicated by the heterogeneity of projects that are studied in online settings (which include creating software, music, and videos and writing encyclopedia entries). Consequently, it is important to ensure that the basis for comparison is limited to the factors that have an impact on collaboration and is not driven by differences between projects. The controlled nature of this field experiment (newly formed teams without prior interaction histories, identical task, identical start and end date, and identical work environment) significantly reduces the risk of team-level unobservables, such as variation in projects (Brock & Durlauf, 2001).

Second, objective performance measures are often not available in realistic work contexts. Here, we have the opportunity to observe teams engaged in a relevant and realistic algorithm development task with the ability to measure the performance of the solutions developed by each team with an objective performance metric.

Third, because participation in many online projects is non-random, important considerations, such as the effect of incentives on participation and individual ability and skills on project performance, need to be factored into the study design. The random assignment of individuals to teams employed in our design implies that we avoid undesirable confounding effects due to sorting and self-selection of individuals into teams. Furthermore, we expect no member interaction outside the team workspace, given the distributed online setting as well as the established norms on the online platform that prohibit interactions across teams.

RESULTS

We begin by examining the effects of our exogenously varied inputs (member skill and cash incentive) along with emergent team collaboration on performance. Our results provide evidence that support the assertion that crowd-based work can be improved by encouraging more and higher quality collaboration among crowd workers, even when controlling for the task-based skills of contributors. Cash incentives have no significant effect on performance, once accounting for member skill and collaboration, in large part because they do not seem to increase collaborative behavior in teams. We will start by discussing the results related to skill and cash incentives and then describe the observations based on the measures of team collaboration that we have devised here: information diversity and burstiness.

Effect of Team-Based Cash Incentives, Member Skill, Information Diversity, and Burstiness on Team Performance

We first examine the first-order effects of exogenous team skill and team-based cash incentives on team performance (Table 4). Based on existing analyses of the effects of cash incentives, we expect an effect of $r = .48$ (Condly, Clark, & Stolovitch, 2003). With a sample size of 52 observed teams, our power to reject the null hypothesis with a target confidence level of 5 percent that cash incentives have an effect of the expected size is .96. Consequently, if we do not find significant effects of cash incentives, this would not be due to the sample size of our study but, rather, would be more likely because of a significantly smaller effect size associated with cash incentives in

TABLE 4
Tobit Regression of Team Score on Team Inputs and Process Measures

Dependent Variable	Team Score						
	Exogenous			Controls	Process Measures		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Randomized							
Team skill	5479.30** (1713.64)		4916.15** (1714.72)	1595.12 (1459.05)	1538.15 (1360.88)	946.31 (1320.43)	1037.49 (1304.02)
Group cash incentive		2496.30* (1234.12)	1695.08 (1156.41)	246.99 (931.38)	570.16 (875.28)	286.61 (832.39)	420.90 (829.50)
Process measures							
Information diversity (standardized)					1526.92** (564.08)		673.06 (614.63)
Burstiness (z-score)						1209.96*** (338.54)	991.61** (383.24)
Controls							
Collective team output				171.21*** (37.57)	116.98** (39.57)	100.39** (37.91)	89.19* (38.67)
Num communications				4.71 (16.78)	−9.15 (16.39)	−32.18 (18.09)	−31.64 (17.80)
Time zone index				−79.21 (270.75)	−90.49 (252.77)	−41.63 (242.95)	−52.77 (239.75)
Num countries				23.11 (536.59)	64.58 (499.54)	531.29 (506.29)	457.15 (501.64)
Constant	3645.76*** (586.86)	2282.87** (871.20)	2841.32*** (811.31)	1698.82 (2383.30)	2302.11 (2231.78)	−1762.39 (2446.92)	−868.77 (2519.36)
Observations	52	52	52	52	52	52	52
Log likelihood	−373.30	−376.22	−372.24	−360.60	−357.07	−354.30	−353.71
Deviance	65.06	65.86	64.50	62.62	61.08	58.83	58.76
Wald test	10.22	4.09	12.52	47.52	59.46	66.31	69.11

***$p < .001$
**$p < .01$
*$p < .05$

our crowdsourcing setting compared with the effect size of cash incentives in traditional team settings.

Model 1 begins by simply regressing our outcome measure, *Team Score*, on our randomized measure *Team Skill*, without any additional controls. We find a significant and positive first-order effect of skill such that teams of higher skill achieve higher scores ($\beta = 5,479.30$; $p < .01$). Model 2 tests a first-order effect of the randomly assigned team-based cash incentives, *Group Cash Incentive*. We find a positive and statistically significant first-order effect indicating that teams under the cash incentive treatment achieve higher scores ($\beta = 2,496.30$; $p < 0.5$). Model 3 uses both the *Team Skill* and *Group Cash Incentive* measures together, which finds only the *Team Skill* term statistically significant ($p < .01$).[5]

We continue our analyses by adding control measures in Model 4. The presence of a cash incentive (*Group Cash Incentive*) has no significant effects on team performance ($\beta = 246.99$; *n.s.*). We find no evidence of team performance being driven by the skill of team members once we include control variables ($\beta = 1,595.12$; *n.s.*). Furthermore, we find no evidence of either positive or negative effects of cultural or time zone diversity in team composition (the coefficients for both *Time Zone Index* and *Num Countries* are not different from zero) and no statistically significant effect of the number of team communications ($\beta = 4.71$; *n.s*). Of the control measures, only the measure of the number of code submissions is statistically significant with a positive coefficient ($\beta = 171.21$; $p < .001$).

Model 5 introduces our measure of *Information Diversity*, which is statistically significant with a positive coefficient ($\beta = 1,526.92$; $p < .01$). The standardized form of the measure allows easy interpretation: A one unit increase in *Information Diversity* corresponds to an increase in team score by about 1,526.92 points—an increase in team score by >37 percent over the mean score. It is important to note that the regression controls for the total number of messages sent within a team, thus controlling for total amount of communication within a team. Consequently, keeping the volume of messages exchanged within a team constant, if those messages contain more diverse information the team performs better. Member skill rating and the presence of cash incentives remain nonsignificant.

Author's voice:
How did you get access to your data or site?

We continue by investigating the burstiness of events within teams as a predictor of team performance. Model 6 regresses our measure of *Burstiness* on team score, including the same vector of controls as before. The regression coefficient of *Burstiness* is statistically significant and positive ($\beta = 1,209.96$; $p < .001$). This indicates that a one standard deviation increase in *burstiness* corresponds to a 1,210 points increase in team score—a 29 percent increase in team performance over the average team. In other words, in the context of this study, the more temporally correlated team activity is—i.e., if there are bursts of high activity—the better a team performs.

Model 7 includes both process measures. We first test the null hypothesis that team processes do not matter for team performance. We perform a likelihood-ratio test to compare model fit of the model with only randomized team inputs and controls (Model 4) and a model which also accounts for emergent collaboration processes (Model 7). This hypothesis is not supported, and there is strong evidence that emergent team processes matter significantly for team performance (log-likelihood ratio: 13.784; $df = 2$; $p < .001$). Next, we investigate the coefficients of the process measures. The measure of *Information Diversity* is not statistically significant ($\beta = 673.06$; *n.s.*) whereas *Burstiness* is statistically significant with a coefficient of comparable size with that of the previous model ($\beta = 991.61$; $p < .01$).

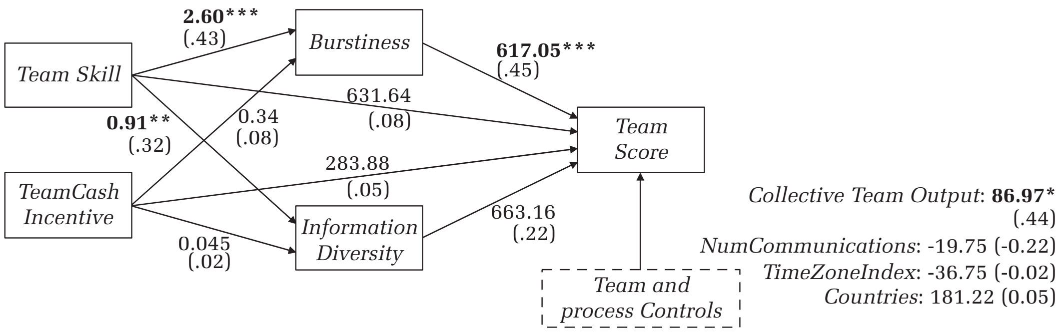
Mediation Analyses: Mechanisms that Link Incentives and Team Skill to Team Performance

The analysis presented previously investigates the effect of exogenous team skill and cash incentives on team performance, whereas simultaneously testing for effects of the emergent team collaboration process measures burstiness and information diversity. It is also possible, however, that team skill and cash incentives have no direct but, rather, an indirect effect on team performance; it is plausible that increased team skill and the presence of cash incentives would lead to improved team processes, which, in turn,

Author's voice:
What was the most difficult or challenging aspect of this research project and paper?

[5] We also test for a significant interaction of *Team Skill* and *Group Cash Incentive* on Team Performance as outcome and both process measures. We find no significant interaction effects.

FIGURE 5
Path Model of Mediation Analysis



Note. Model reports unstandardized regression coefficients and significance levels based on structural equation modeling with 1,000 bootstrap simulations. Standardized regression coefficients in parentheses.
***$p < 0.001$
**$p < 0.01$
*$p < 0.05$

translate to higher team performance. Thus, we are interested in testing whether there are significant *indirect* paths (i.e., mediation effects) that link team skill and cash incentives to team performance via the mechanisms of burstiness and information diversity.

We use a structural equation model (SEM) to test whether *Burstiness* and *Information Diversity* mediate the effect of team skill and cash incentives on team performance. SEM has several advantages in testing mediation effects compared with the hierarchical regression approach (Cheung & Lau, 2008). Specifically, SEM allows testing of multiple mediators simultaneously and has been shown to be more robust in the presence of measurement error. We perform SEM using maximum-likelihood estimators (Bollen, 2005), which we carried out using the lavaan package (Rosseel, 2012) for structural equation modeling implemented for R.

Similar to the hierarchical regression approach, bootstrapping can be used instead of distribution-based tests, such as the Sobel test in SEM, to test for statistical significance and compute confidence intervals (CIs) of indirect effects. We use Bollen-Stine's model-based bootstrapping (1,000 simulations) to test statistical significance and the adjusted bootstrap percentile (BC) method to construct confidence intervals. Despite some general concerns using mediation-analysis on small samples (Koopman, Howe, Hollenbeck, & Sin, 2015), bootstrapping (especially the corrected BC method that we use) has been shown to be superior to other approaches of testing mediation (Cheung & Lau, 2008; MacKinnon, Fairchild, & Fritz, 2007). The results of the structural model testing mediation are shown in Figure 5 (χ^2 = 100.32; p < .001; BIC = 2,682.56). The path from *Team Skill* to *Burstiness* is statistically significant ($\beta = 2.6$; $p < .001$) as is the path to *Information Diversity* ($\beta = .91$; $p < .01$). The paths from *Team Cash Incentive* to *Burstiness* and *Information Diversity* are not significant. The path from *Burstiness* to *Team Score* is statistically significant ($\beta = 617.05$; $p < .001$), the path from *Information Diversity* to *Team Score* is not ($\beta = 663.16$; *n.s.*). Simultaneously, we find no statistically significant direct effect of *Team Skill* or *Team Cash Incentives* on *Team Score*. This indicates that the two process measures *Burstiness* and *Information Diversity* together fully mediate the effects of exogenous team inputs on team performance.

Next, we compute confidence intervals for the indirect effects. In terms of unstandardized regression coefficients, we find an indirect effect via *Burstiness* of 1,816.62 [95% CI: 714.88 to 3,733.61] ($p < .05$), an indirect effect via *Information Diversity* of 634.40 [95% CI: −65.08 to 1,624.84] ($p > .05$), a combined indirect effect of 2,451.02 [95% CI: 939.79 to 3,704.31] ($p < .001$), and a total effect of 3491.25 [95% CI: 949.27 to 5,755.72] ($p < .01$).

Given the nature of our study design, which allows us to estimate causal effects, we also are specifically interested in identifying causal mechanisms. Recent advances in the development of statistical methods have demonstrated that the linear structural equation model is nonparametrically identified, and estimates can be interpreted as average causal mediation effect (ACME) under certain assumptions (Imai, Keele, & Yamamoto, 2010). We perform robustness tests that employ these new methods to examine whether the mediation effects described previously can be interpreted causally. Specifically, we carried out causal mediation analysis using the mediation package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) for R. We find a statistically significant ACME of *Burstiness* (ACME = 2,301.87 [95% CI: 927.92 to 5,106.40]; $p < .001$). Contrary to the SEM analysis, we also find a statistically significant ACME of *Information Diversity* (ACME = 1,016.109 [95% CI: 277.64 to 2,520.53]; $p < .001$). In summary, analyses that use these alternative methods find that both process measures mediate the relationship between exogenous team skill and team performance and suggest that this mediation can be interpreted causally.

DISCUSSION AND CONCLUSION

In this study, we experimented with implementing team structures in a typical crowdsourcing contest platform and found that exogenous team inputs (member skill and cash incentives) were influential for performance only if they were translated into effective team collaboration. Our empirical analysis of our field experiment indicates (1) a moderate effect for member skill and a weak effect for cash incentives on team performance; and (2) large, positive effects of emergent collaboration on team performance, specifically related to the burstiness of team activities and information diversity, even when the effects of member skill and cash incentives are controlled. Our statistical results indicate particularly large effects associated with burstiness. These effects are especially noteworthy, given that these crowd-based teams were distributed across the world and participating outside of "normal" organizational roles. Yet, we see that synchrony dynamics that one would expect to be important only to traditional face-to-face collaborations still prevail.

Our findings have significant implications for how to best structure and manage crowdsourced problem solving, as well as teamwork in traditional settings. In the following sections, we discuss implications for crowdsourcing research, team research, and how our methodology for measuring burstiness and information diversity can enable the examination of team process in a wider variety of environments.

Burstiness and Information Diversity

Our study suggests that informational diversity and burstiness can be readily observed in teams and studied based on time-stamped archival information. We observed that high performing teams had fewer topic repetitions in their communication and, rather, drew on a more diverse set of topics. Furthermore, we found that higher performing teams were more "bursty," as they coordinated their activities such that at least some messages received rapid responses, for example, through appropriate prioritization. This is striking in this context, particularly given the time zone differences spanned by the teams.

Although our findings related to burstiness in teams contribute to a developing literature on temporal dynamics in team performance, they are, at the same time, distinct from existing work. For instance, increasing attention has been paid to the use of temporal milestones and the pacing of work in teams (Gersick, 1988, 1989, 1991) and the use of time as a semi-structure for evaluating progress and coordinating work (Okhuysen & Waller, 2002). The punctuation in that work relates specifically to *change* and transition; groups in related studies often met regularly, in some cases continuously, throughout their work. By contrast, the work presented here documents bursts in all team-related activity, regardless of whether those activities marked changes and transitions or just regular task work. Furthermore, we do not observe a pattern in this study that parallels those observed in the punctuated equilibrium work as bursts in the teams observed here happened throughout the work period versus exhibiting concentration at the beginning, midpoint, and/or end of work.

Burstiness in these teams also differs from existing work on temporal entrainment, in which teams align with externally imposed rhythms that constrain their routines (Zellmer-Bruhn et al., 2004). In the current study, we do not observe evidence of the teams' entraining to any externally or internally imposed rhythm or schedule. Indeed, what is remarkable in the bursty teams in our study is the high level of responsiveness and temporal coordination, despite their being in different time zones, which would lead members to work at wildly varying and unpredictable hours around the clock. In this way, burstiness diverges from concepts of entrainment, in which activities are aligned with external temporal events and predictably timed (Ancona & Chong, 1999).

In summary, we do not view the observations of burstiness in the current study as contradicting existing work on temporal patterns in teams but, instead, as providing a new angle that is particularly applicable to these globally distributed crowd-based collaborations. Members of these teams were participating in these projects in addition to the activities of their "regular" life, and, thus, the ability to draw one another in to collaborate more synchronously is particularly challenging in this setting and, as such, is a bigger differentiator among teams. As we see work in an increasing number of organizations structured around ad hoc projects and multiple team membership (O'Leary, Mortensen, & Woolley, 2011), we may see burstiness become increasingly important there as well.

Incentives and Collaboration in Crowdsourcing

In the current study, although team-based cash incentives had a significant first-order effect on performance, cash incentives had no effect on the level and quality of collaboration observed within teams. Thus, our study has implications for the role that incentive systems play in crowdsourced collaboration settings. The design of incentive mechanisms has been a constant focus in much management and economics research (Boudreau et al., 2011; Hamilton, Nickerson, & Owan, 2003; Prendergast, 1999). Most of this research, however, is focused on the link between incentives and effort rather than on incentives and collaboration. The current study suggests that, for collaborative crowdsourcing, incentives should be accompanied by other tools and structures that encourage collaboration. The size and probability of winning a cash reward in this contest were rather large relative to other contexts; however, future research could also explore if larger incentives might elicit a stronger response.

With regard to crowd-based collaboration, our findings support the recommendation that crowd-based teams engage in some level of temporal coordination of their inputs so that they can achieve the performance benefits of burstiness. Online platforms for crowdsourcing could also integrate technologies that would help teams do this more easily. Cues that allow team members to see other members' activities, such as when they are online or what they are working on, even if they are not communicating, might help facilitate greater burstiness. Other cues that prompt members to think about a wider variety of issues about which to communicate, such as periodic messages or prompts in the team workspace, could help facilitate greater diversity of information sharing. These directions can serve as an important complement to current approaches that seek to automate the coordination of crowd-based teams (Kittur et al., 2011; Retelny et al., 2014). As others have pointed out (Blohm, Riedl, Füller, & Leimeister, 2016; Majchrzak & Malhotra, 2013), given the richness in the technologies that are becoming available to facilitate collaboration, we could look for more ways to encourage the collaboration processes that we know are important for crowdsourced work.

Future Directions

Although our findings contribute to answering a number of questions about enhancing emergent collaboration in online environments, a number of questions for future research are suggested as well. Although participants were informed that they might be working in teams in this contest, most were not accustomed to working in this manner in a contest of this type. Future research could examine how the behavior of those who self-select into team collaboration differs from those who prefer to work individually to gauge how this would change the factors that are most influential to performance. In addition, although participants who were assigned to our "incentive" condition were eligible for a prize, none of our participants were directly paid for their effort. Thus, the degree to which our findings would generalize to a setting in which salaried employees are paid for their time to work on an online platform is a matter that also should be examined in future research.

The questions that our study answers represent an important extension of the literature on crowdsourcing and team performance in a number of ways. First and foremost, the unique nature of this field setting allows us to make stronger causal claims about the effects of cash incentives and collaboration on performance than has been possible in most prior studies. In particular, although the effects of incentives, information diversity, and burstiness have been studied independently in prior work, we are able to examine their independent and combined effects on performance. Second, the completeness of our observed data on the teams' collaboration as well as the high fidelity of our metrics of individual skills and team performance allow us to examine and control for important variables that are often treated as sources of error in more traditional field studies of team collaboration. Thus, we are able to more precisely specify the effects of the team process variables and underscore their importance as a design principle for improving collaboration in crowdsourcing environments.

Our data demonstrate that emergent team collaboration processes are important to team performance in crowdsourcing environments. Our findings suggest ways that online environments could be designed to more consistently bring about a high level of performance. We hope that our findings as well as the measures that we have introduced for assessing emergent collaboration will be helpful in future research on crowd-based as well as traditional teams.

REFERENCES

Afuah, A., & Tucci, C. L. 2012. Crowdsourcing as a solution to distant search. ***Academy of Management Review***, 37(3): 355–375.

Alchian, A. A., & Demsetz, H. 1972. Production, information costs, and economic organization. ***American Economic Review***, 62(5): 777–795.

Amabile, T. M. 1993. Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. ***Human Resource Management Review***, 3(3): 185–201.

Amabile, T. M., & Fisher, C. M. 2009. Stimulate creativity by fueling passion. In E. Locke (Ed.), ***Handbook of principles of organizational behavior: Indispensable knowledge for evidence-based management***: 479–494. West Sussex, UK: John Wiley & Sons.

Ancona, D. G., & Chong, C. 1999. Cycles and synchrony: The temporal role of context in team behavior. In E. Mannix & M. A. Neale (Eds.), ***Research on managing groups and teams***, vol. 2. Greenwich, CT: JAI Press.

Aral, S., & Van Alstyne, M. 2011. The diversity-bandwidth tradeoff. ***American Journal of Sociology***, 117(1): 90–171.

Arolas, E. E., & Ladron-de-Guevara, F. G. 2012. Towards an integrated crowdsourcing definition. ***Journal of Information Science***, 38(2): 189–200.

Austin, J. R. 1997. A cognitive framework for understanding demographic influences in groups. ***International Journal of Organizational Analysis***, 5(4): 342–359.

Bantel, K. A., & Jackson, S. E. 1989. Top management and innovations in banking: Does the composition of the top team make a difference? ***Strategic Management Journal***, 10(S1): 107–124.

Barabasi, A.-L. 2005. The origin of bursts and heavy tails in human dynamics. ***Nature***, 435(7039): 207–211.

Barabási, A.-L. 2010. ***Bursts: The hidden patterns behind everything we do, from your e-mail to bloody crusades***. New York: Penguin.

Bell, S. T. 2007. Deep-level composition variables as predictors of team performance: A meta-analysis. ***Journal of Applied Psychology***, 92(3): 595–615.

Blei, D., Ng, A., & Jordan, M. 2003. Latent dirichlet allocation. ***Journal of Machine Learning Research***, 3(January): 993–1022.

Blohm, I., Riedl, C., Füller, J., & Leimeister, J. M. 2016. Rate or trade? Identifying winning ideas in open idea sourcing. ***Information Systems Research***, 27(1): 27–48.

Bollen, K. A. 2005. ***Structural equation models***. New York: Wiley Online Library.

Boudreau, K. J. 2012. Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. ***Organization Science***, 23 (5): 1409–1427.

Boudreau, K. J., Lacetera, N., & Lakhani, K. R. 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. ***Management Science***, 57(5): 843–863.

Boudreau, K. J., & Lakhani, K. R. 2013. Using the crowd as an innovation partner. ***Harvard Business Review***, 91 (4): 60–69.

Brock, W. A., & Durlauf, S. N. 2001. Interactions-based models. ***Handbook of Econometrics***, 5: 3297–3380.

Cacioppo, S., & Cacioppo, J. T. 2012. Decoding the invisible forces of social connections. ***Frontiers in Integrative Neuroscience***, 6.

Caruso, H. M., & Woolley, A. W. 2008. Harnessing the power of emergent interdependence to promote diverse team collaboration. In K. W. Phillips, E. Mannix, & M. A. Neale (Eds.), vol. 11: 245–266. Bingley, UK: Emerald Group Publishing Limited.

Chatman, J. A., Polzer, J. T., Barsade, S. G., & Neale, M. A. 1998. Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. ***Administrative Science Quarterly***, 43(4): 749–780.

Cheung, G. W., & Lau, R. S. 2008. Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. ***Organizational Research Methods***, 11(2): 296–325.

Chikersal, P., Tomprou, M., Kim, Y. J., Woolley, A., & Dabbish, L. 2017. Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction. Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing in Portland, OR. New York: ACM. pp. 873–888.

Condly, S. J., Clark, R. E., & Stolovitch, H. D. 2003. The effects of incentives on workplace performance: A meta-analytic review of research studies. ***Performance Improvement Quarterly***, 16(3): 46–63.

Cramton, C. D. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. ***Organization Science***, 12(3): 346–71.

Cronin, M. A., & Weingart, L. R. 2007. Representational gaps, information processing, and conflict in functionally diverse teams. ***Academy of Management Review***, 32(3): 761–773.

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

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. 2005. Team diversity and information use. ***Academy of Management Journal***, 48(6): 1107–1123.

De Dreu, C. K., & West, M. A. 2001. Minority dissent and team innovation: The importance of participation in decision making. ***Journal of Applied Psychology***, 86 (6): 1191–1201.

Deci, E. L., & Ryan, R. M. 1985. ***Intrinsic motivation and self-determination in human behavior***. New York: Plenum Press.

Devine, D. J. 1999. Effects of cognitive ability, task knowledge, information sharing, and conflict on group decision-making effectiveness. ***Small Group Research***, 30(5): 608–634.

Devine, D. J., & Philips, J. L. 2001. Do smarter teams do better? A meta-analysis of cognitive ability and team performance. ***Small Group Research***, 32(5): 507–532.

Du Chatenier, E., Verstegen, J. A. A. M., Biemans, H. J. A., Mulder, M., & Omta, O. 2009. The challenges of collaborative knowledge creation in open innovation teams. ***Human Resource Development Review***, 8(3): 350–381.

Frey, B. S., & Jegen, R. 2001. Motivation crowding theory. ***Journal of Economic Surveys***, 15(5): 589–611.

Fullerton, R., Linster, B. G., McKee, M., & Slate, S. 1999. An experimental investigation of research tournaments. ***Economic Inquiry***, 37(4): 624–636.

Gersick, C. J. G. 1988. Time and transition in work teams: Toward a new model of group development. ***Academy of Management Journal***, 31: 9–41.

Gersick, C. J. G. 1989. Marking time: Predictable transitions in task groups. ***Academy of Management Journal***, 32 (2): 274–309.

Gersick, C. J. G. 1991. Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. ***Academy of Management Review***, 16: 10–36.

Gibson, C. B., & Gibbs, J. L. 2006. Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. ***Administrative Science Quarterly***, 51(3): 451–495.

Gladstein, D. L. 1984. Groups in context: A model of task group effectiveness. ***Administrative Science Quarterly***, 29(4): 499–517.

Goh, K.-I., & Barabási, A.-L. 2008. Burstiness and memory in complex systems. ***EPL (Europhysics Letters)***, 81(4): 48002.

Hackman, J. R. 1987. ***The design of work teams***. In J. W. Lorsch (Ed.), ***Handbook of organizational behavior***: 315–342. Englewood Cliffs, NJ: Prentice Hall.

Hackman, J. R. 2011. ***Collaborative intelligence: Using teams to solve hard problems***. San Francisco, CA: Berrett-Koehler Publishers.

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

Harrison, D. D., & 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.

Harrison, D., Price, K., & Bell, M. 1998. Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. ***Academy of Management Journal***, 41: 96–107.

Howe, J. 2006. The rise of crowdsourcing. ***WIRED***, (June). http://www.wired.com/2006/06/crowds/. Accessed December 12, 2017.

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(1): 517–543.

Imai, K., Keele, L., & Yamamoto, T. 2010. Identification, inference, and sensitivity analysis for causal mediation effects. ***Statistical Science***, 25(1): 51–71.

Janicik, G. A., & Bartel, C. A. 2003. Talking about time: Effects of temporal planning and time awareness norms on group coordination and performance. ***Group Dynamics: Theory, Research, and Practice***, 7 (2): 122–134.

Jenkins, G. D., Mitra, A., Gupta, N., & Shaw, J. D. 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. ***Journal of Applied Psychology***, 83(5): 777.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and problem-solving effectiveness in broadcast search. ***Organization Science***, 21(5): 1016–1033.

Karsai, M., Kaski, K., Barabási, A.-L., & Kertész, J. 2012. Universal features of correlated bursty behaviour. ***Scientific Reports***, 2: 397.

Kilduff, M., Angelmar, R., & Mehra, A. 2000. Top management-team diversity and firm performance: Examining the role of cognitions. ***Organization Science***, 11(1): 21.

Kittur, A., Nickerson, J. V., Bernstein, M. S., Gerber, E. M., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. J. 2013. The future of crowd work. Proceedings of the 2013 Conference on Computer Supported Cooperative Work, 1301–1318. San Antonio, TX: ACM.

Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. 2011. CrowdForge: Crowdsourcing complex work. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, 43–52. New York: ACM.

Kleiber, C., & Zeileis, A. 2008. ***Applied econometrics with R***. New York: Springer-Verlag.

Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. 2015. Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. ***Journal of Applied Psychology***, 100(1): 194–202.

Kozlowski, S. W. J. 2015. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. ***Organizational Psychology Review***, 5(4): 270–299.

Krych-Appelbaum, M., Law, J. B., Jones, D., Barnacz, A., Johnson, A., & Keenan, J. P. 2007. "I think I know what you mean": The role of theory of mind in collaborative communication. ***Interaction Studies***, 8(2): 267–280.

Lazear, E. P. 2000. Performance pay and productivity. ***American Economic Review***, 90(5): 1346–1361.

Lazear, E. P., & Shaw, K. L. 2007. Personnel economics: The Economist's view of human resources. ***Journal of Economic Perspectives***, 21(4): 91–114.

Lazer, D., & Friedman, A. 2007. The network structure of exploration and exploitation. ***Administrative Science Quarterly***, 52(4): 667–694.

LePine, J. A. 2003. Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. ***Journal of Applied Psychology***, 88(1): 27.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. 2007. Mediation analysis. ***Annual Review of Psychology***, 58(1): 593–614.

Majchrzak, A., & Malhotra, A. 2013. Towards an information systems perspective and research agenda on crowdsourcing for innovation. ***Journal of Strategic Information Systems***, 22(4): 257–268.

Malmgren, R. D., Stouffer, D. B., Motter, A. E., & Amaral, L. A. N. 2008. A poissonian explanation for heavy tails in e-mail communication. ***Proceedings of the National Academy of Sciences***, 105(47): 18153–18158.

Malone, T. W., Laubacher, R., & Dellarocas, C. 2010. The collective intelligence genome. ***MIT Sloan Management Review***, 51(3): 21–31.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. 2001. A temporally based framework and taxonomy of team processes. ***Academy of Management Review***, 26(3): 356–376.

Massey, A. P., Montoya-Weiss, M. M., & Hung, Y.-T. 2003. Because time matters: Temporal coordination in global virtual project teams. ***Journal of Management Information Systems***, 19(4): 129–155.

McGrath, J. E. 1991. Time, interaction, and performance (TIP): A theory of groups. ***Small Group Research***, 22 (2): 147–174.

McLeod, P. L., Lobel, S. A., & Cox, T. H. 1996. Ethnic diversity and creativity in small groups. ***Small Group Research***, 27(2): 248–264.

Mello, A. L., & Rentsch, J. R. 2015. Cognitive diversity in teams a multidisciplinary review. ***Small Group Research***, 46(6): 623–658.

Milliken, F. J., Bartel, C. A., & Kurtzberg, T. R. 2003. Diversity and creativity in work groups: A dynamic perspective on the affective and cognitive processes that link diversity and performance. In Paulus, P. B. & Nijstad, B. A. ***Group Creativity: Innovation Through Collaboration***. New York: Oxford University Press.

Milliken, F. J., & Martins, L. L. 1996. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. ***Academy of Management Review***, 21(2): 402–433.

Mohammed, S., & Ringseis, E. 2001. Cognitive diversity and consensus in group decision making: The role of inputs, processes, and outcomes. ***Organizational Behavior and Human Decision Processes***, 85(2): 310–335.

Moldovanu, B., & Sela, A. 2006. Contest architecture. ***Journal of Economic Theory***, 126(1): 70–96.

Montoya-Weiss, M. M., Massey, A. P., & Song, M. 2001. Getting it together: Temporal coordination and conflict management in global virtual teams. ***Academy of Management Journal***, 1251–1262.

Nalbantian, H. R., & Schotter, A. 1997. Productivity under group incentives: An experimental study. ***American Economic Review***, 87(3): 314–341.

Ocker, R., Hiltz, S. R., Turoff, M., & Fjermestad, J. 1995. The effects of distributed group support and process structuring on software requirements development teams: Results on creativity and quality. ***Journal of Management Information Systems***, 12(3): 127–153.

Okhuysen, G. A., & Waller, M. J. 2002. Focusing on midpoint transitions: An analysis of boundary conditions. ***Academy of Management Journal***, 45(5): 1056–1065.

O'Leary, M. B., & Cummings, J. N. 2007. The spatial, temporal, and configurational characteristics of geographic dispersion in work teams. ***MIS Quarterly***, 31 (3): 433–452.

O'Leary, M. B., & Mortensen, M. 2010. Go (con)figure: Subgroups, imbalance, and isolates in geographically dispersed teams. ***Organization Science***, 21(1): 115–131.

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

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., & deVreede, G. J. 2013. Conceptual foundations of crowdsourcing: A review of IS research, 579–588. In proceedings of the System Sciences, 2013 46th Hawaii International Conference, Maui, HI: IEEE.

Phillips, K. W., & Loyd, D. L. 2006. When surface and deep-level diversity collide: The effects on dissenting group members. ***Organizational Behavior and Human Decision Processes***, 99(2): 143–160.

Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. 2004. Diverse groups and information sharing: The effects of congruent ties. ***Journal of Experimental Social Psychology***, 40(4): 497–510.

Prendergast, C. 1999. The provision of incentives in firms. ***Journal of Economic Literature***, 37(1): 7–63.

R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/. Accessed December 12, 2017.

Reagans, R., Zuckerman, E., & McEvily, B. 2004. How to make the team: Social networks vs. demography as criteria for designing effective teams. ***Administrative Science Quarterly***, 49(1): 101–133.

Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., Doshi, T., Valentine, M., & Bernstein, M. S. 2014. ***Expert crowdsourcing with flash teams***. Proceedings of the 27th annual ACM symposium on User interface software and technology, 75–85. Honolulu, HI: ACM.

Rosseel, Y. 2012. Lavaan: An R package for structural equation modeling. ***Journal of Statistical Software***, 48(2): 1–36.

Stasser, G., Stewart, D. D., & Wittenbaum, G. M. 1995. Expert roles and information exchange during discussion: The importance of knowing who knows what. ***Journal of Experimental Social Psychology***, 31: 244–265.

Stephan, P. E. 1996. The economics of science. ***Journal of Economic Literature***, 34(3): 1199–1235.

Surowiecki, J. 2004. ***The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations***. New York: Anchor Books.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. 2014. mediation: R package for causal mediation analysis. ***Journal of Statistical Software***, 59(5): 1–38.

Wageman, R. 1995. Interdependence and group effectiveness. ***Administrative Science Quarterly***, 40(1): 145–180.

Wageman, R., & Baker, G. 1997. Incentives and cooperation: The joint effects of task and reward interdependence on group performance. ***Journal of Organizational Behavior***, 18(2): 139–158.

Wageman, R., & Gordon, F. M. 2005. As the twig is bent: How group values shape emergent task interdependence in groups. ***Organization Science***, 16(6): 687–700.

Warkentin, M. E., Sayeed, L., & Hightower, R. 1997. Virtual teams versus face-to-face teams: An exploratory study of a web-based conference system. ***Decision Sciences***, 28(4): 975–996.

Williams, K. Y., & O'Reilly, C. A. I. 1998. Demography and diversity in organizations: A review of 40 years of research. In L. L. Cummings (Ed.), ***Research in organizational behavior***, vol. 20: 77–140. Greenwich, CT: JAI Press.

Wiltermuth, S. S., & Heath, C. 2009. Synchrony and cooperation. ***Psychological Science***, 20(1): 1–5.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. 2010. Evidence for a collective intelligence factor in the performance of human groups. ***Science***, 330(6004): 686–688.

Wooten, J. O., & Ulrich, K. T. (2017). Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments. ***Production and Operations Management***, 26(1), 80–99.

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

Zellmer-Bruhn, M., Waller, M. J., & Ancona, D. G. 2004. The effect of temporal entrainment on the ability of teams to change their routines. ***Research on Managing Groups and Teams***, 6: 135–138.

Zenger, T. R., & Lawrence, B. S. 1989. Organizational demography: The differential effects of age and tenure distributions on technical communication. ***Academy of Management Journal***, 32(2): 353–376.

Christoph Riedl (c.riedl@northeastern.edu) is an assistant professor of Information Systems at the D'Amore-McKim School of Business and the College of Computer and Information Science at Northeastern University. He is a fellow at the Institute for Quantitative Social Science at Harvard University. He obtained his PhD from Technische Universität München. His research focuses on collaborative problem-solving and decision-making.

Anita Williams Woolley (awoolley@cmu.edu) is an associate professor of Organizational Behavior and Theory at the Tepper School of Business, Carnegie Mellon University and an affiliate of the Center for Collective Intelligence at the Massachusetts Institute of Technology. She obtained her PhD from Harvard University. Her research focuses on collective intelligence and collaboration in teams.