

Aligning Differences: Discursive Diversity and Team Performance

Katharina Lix

Stanford University, klix@stanford.edu

Amir Goldberg

Stanford University, amirgo@stanford.edu

Sameer B. Srivastava

University of California, Berkeley, sameersriv@berkeley.edu

Melissa A. Valentine

Stanford University, mav@stanford.edu

How does cognitive diversity in a group affect its performance? Prior research suggests that group cognitive diversity poses a performance tradeoff: diverse groups excel at creativity and innovation but struggle to take coordinated action. Building on the insight that group cognition is not static but is instead dynamically and interactively produced, we introduce the construct of *discursive diversity*, a manifestation of group cognitive diversity, which reflects the degree to which the meanings conveyed by group members in a given set of interactions diverge from one another. We propose that high-performing teams are ones that have a collective capacity to modulate shared cognition to match changing task requirements: they exhibit higher discursive diversity when engaged in ideational tasks and lower discursive diversity when performing coordination tasks. We further argue that teams exhibiting *congruent modulation*—that is, those with low group-level variance in members’ within-person semantic shifts to changing task requirements—are more likely to experience success than teams characterized by incongruent modulation. Using the tools of computational linguistics to derive a measure of discursive diversity and drawing on a novel longitudinal data set of intragroup electronic communications and performance outcomes for 117 remote software development teams on an online platform (www.gigster.com), we find support for our theory. Our findings suggest that the performance tradeoff of group cognitive diversity is not inescapable: groups can navigate it by aligning their levels of discursive diversity to match their task requirements and by having members stay aligned with one another as they make these adjustments.

Key words: groups and teams, cognition, diversity, interaction

Introduction

Why do some groups perform better than others when working toward a shared goal? An extensive literature has examined this question through the lens of group cognitive diversity. The prevailing view, backed by a substantial body of empirical evidence, posits that cognitive diversity embodies a performance tradeoff: diverse groups draw on a broader set of ideas and are therefore better at discovering novel and effective solutions (e.g., Page 2008, Gibson and Vermeulen 2003), but this collective problem-solving ability comes at the expense of coordinated action, which is easier to achieve when group members’ interpretations are aligned (e.g., Sørensen 2002, March 1991, Knight et al. 1999).

Scholars have uncovered this tension in a variety of contexts and at different levels of analysis. For example, cognitive diversity in a firm’s executive team inhibited thorough examinations of threats and opportunities, as well as long-range planning (Miller et al. 1998). In contrast, cognitive diversity in MBA teams provided members with signals about where to locate critical resources needed to support creative production (Aggarwal and Woolley 2019). And firms whose members held a breadth of interpretations about their organizational culture were better at creative innovation, while those whose members had clashing interpretations struggled to coordinate effectively and were less profitable (Corritore et al. 2020).

The tradeoffs of convergent versus divergent thinking for group performance have been extensively studied in work on shared cognition in teams. When team members approach problems from different perspectives, they can collectively develop novel insights that no individual could have conceived of independently (Pelled et al. 1999, Amabile et al. 1996, Aggarwal and Woolley 2019). At the same time, teams can perform at a high level when each contributor understands and approaches tasks in a consistent manner, thereby enabling better communication and smoother coordination (Converse et al. 1993, Cropley 2006).

Although various contingencies—such as the network structure of teams (e.g., Reagans and Zuckerman 2001, Reagans et al. 2004)—have been identified, the diversity-performance tradeoff has generally been thought to be inescapable. We argue that this conclusion is based on the assumption that the ideas people bring to a group, and the behaviors these ideas catalyze, are mostly predetermined and relatively stable. Yet a large body of work by interactional sociologists and social psychologists demonstrates that people produce meaning dynamically through group interaction (Berger and Luckmann 1967, Eliasoph and Lichterman 2003, Knorr Cetina and Bruegger 2002). Consistent with this view, a nascent but growing literature suggests that groups often elude simple categorization as diverse or homogeneous; rather, group members interact in ways that at times surface and amplify divergent ideas and in other instances smooth and dampen these differences (Cronin et al. 2011, Srikanth et al. 2016).

Building on this fundamental understanding of group cognition as dynamically and interactively produced, we argue that teams’ shared cognition can fluctuate between states of convergence and divergence and that these shifts have implications for the performance of the group as a whole. In developing this argument, we introduce the construct of discursive diversity, a manifestation of group cognitive diversity, which reflects the degree to which the meanings conveyed by group members in a given set of interactions diverge from one another. We propose that high-performing teams are characterized by a collective capability to modulate shared cognition to match their changing task requirements: they exhibit higher levels of discursive diversity when engaged in ideational tasks and lower levels of discursive diversity when performing coordination tasks. We further argue that teams exhibiting *congruent modulation*—those with lower group-level variance in members’ within-person semantic shifts to changing task requirements—are more likely to experience success relative to teams characterized by incongruent modulation.

Prior work has assessed shared cognition in groups using self-reports, which yield static, or at best episodic, measures and are known to suffer from various forms of reporting bias (Greenwald and Banaji 1995, Srivastava and Banaji 2011). As an alternative, we propose that the natural language through which group members communicate with one another can be harnessed to yield fine-grained, behavioral measures of group cognitive diversity.

Building on recent work that uses language as a window into different facets of group cultural dynamics (Goldberg et al. 2016, Srivastava et al. 2018), we harness the tools of computational linguistics to derive a language-based measure of discursive diversity. Applying this measure to a longitudinal data set of intragroup electronic communications and team performance outcomes for 117 teams on an online platform (www.gigster.com) that matches freelance software developers and project managers to work projects, we find support for our theoretical predictions. We discuss implications for research on group shared cognition, collective meaning making, and organizational ambidexterity.

Theory

Shared Cognition in Teams

Teams have long been recognized as social information-processing units (Griffith and Neale 2001, Hackman 2011, Salancik and Pfeffer 1978). Extensive research shows that team outputs often exceed the sum of individual members’ contributions (Larson 2010, Wuchty et al. 2007). These synergistic group effects arise from social facilitation (e.g., Guerin 1986, Harkins 1987), specialization (e.g., Bunderson and Boumgarden 2010, Lewis 2002), and the novel recombination of information (e.g., Cummings 2004, De Dreu and West 2001). A large literature has explored the interactional processes that enable groups to produce excess value, and many group-level properties—including

collective intelligence (e.g., Woolley et al. 2010), shared beliefs (e.g., Edmondson 1999), and shared cognition (e.g., Carley 1997)—have been conceptualized and measured.

Within this overall research agenda, considerable attention has been paid to the role of shared cognition in shaping team performance. Shared cognition broadly refers to “how dyads, groups, and larger collectives create and utilize interpersonal understanding” (Thompson and Fine 1999, p. 280). Prior research has identified several facets of group shared cognition. A prominent stream of work has, for example, focused on team mental models: “organized knowledge structures that allow individuals to interact with their environment” (Mathieu et al. 2000, p.276). Another line of research has examined transactive memory systems—the cognitive architecture that encompasses both the knowledge uniquely held by particular group members and the collective awareness of who knows what (Liang et al. 1995, Moreland et al. 1998, 1996). Examples of other shared cognition constructs include strategic consensus, collective mind, and cognitive style diversity (DeChurch and Mesmer-Magnus 2010, Mohammed et al. 2010).

These different approaches to shared cognition are unified in assuming the existence of latent knowledge structures in individuals’ minds that, to varying degrees, are shared across team members. Given the limitations of collecting longitudinal data on group processes via self-reports, prior work has tended to conceptualize shared cognition as a group property that is either stable or that shifts at most periodically.

Discursive Diversity as a Manifestation of Group Cognitive Diversity

We propose a different approach. Following Weick (1979, p.5)—who asked, “How can I know what I think until I see what I say?”—we assume that shared cognition and discourse are inherently intertwined.¹ Instead of examining shared cognition by probing team members’ latent knowledge structures, we examine the structures of meanings that emerge through their discursive interaction.

Shared meaning emerges through interaction because people adjust their interpretations of a situation in response to the meanings expressed by others (Healey et al. 2015, Thompson and Fine 1999). Thus, group cognitive diversity is reflected in the extent to which the *discourse* between group members in a given interaction is divergent rather than convergent. By discourse we do not simply mean the set of words expressed in language. Rather, we use discourse to connote the underlying meanings communicated in conversation and the ways in which they reflect interaction partners’ structures of knowledge and interpretation (Foucault 2002). We refer to the diversity in the meanings that team members convey to one another in a given interaction as *discursive diversity*.

¹We thank an anonymous reviewer for suggesting that we reference this quotation in building our conceptual argument.

Whereas existing approaches identify shared cognition by tapping team members’ retrospective and reified perceptions of team processes, we locate it in the discourse that they collectively and interactively produce. This approach, we contend, has several advantages. First, our approach has higher fidelity to the process through which meanings are negotiated in a team setting. In conventional studies of shared cognition, such as those focused on shared mental models, team members are often asked to report which attributes of a group task are germane or how they relate to one another. In real life, however, participants very rarely, if ever, deliberately reflect on a task in such a disciplined and abstract fashion. Rather, as we explain below, our approach induces the meanings that team members associate with their interactive experiences by examining their expressed language.

Second, this inductive approach does not require us to make a priori assumptions about the dimensions of meaning that are relevant for participants. We therefore do not constrain our analysis to the dimensions of cognition that we imagine as relevant to group performance.

Finally, unlike established constructs that emphasize specific knowledge structures—for example, mental models related to tasks (e.g., procedures, contingencies, scenarios, and strategies) or to properties of the team (e.g., respective members’ skills and responsibilities or expected interaction patterns) (Mathieu et al. 2000, p.276)—discursive diversity encompasses the full array of meanings that group members convey to each other. We turn next to considering the consequences of discursive diversity for team performance.

Discursive Diversity, Task Requirements, and Team Performance

As noted above, existing work sees group cognitive diversity as presenting a tradeoff: groups can either innovate and learn by being cognitively divergent, or they can coordinate effectively by being cognitively convergent. In this view, maximizing creativity and innovation generally comes at the expense of coordination effectiveness, and vice versa. We propose that, insofar as groups have the capacity to modulate their levels of discursive diversity to match their shifting task requirements, they can successfully navigate the diversity-performance tradeoff.

Indeed, the match between group processes and the nature of the task being performed is critical for team success.² Group tasks can be broadly categorized into two task types: ideation and coordination (McGrath 1991, Fleishman and Zaccaro 1992, Prince and Salas 1993, Marks et al. 2001). Drawing on this distinction, we posit that high-performing teams exhibit higher discursive diversity when engaged in ideational tasks and lower discursive diversity when performing

²Team success can, of course, be defined along multiple dimensions such as speed, cost efficiency, creativity, and learning. Here we focus on success as defined by timely completion of a team’s project goals, which are typically set and evaluated relative to customer requirements. In the context of the present study, successful goal completion required both coordination and ideation.

coordination tasks. This is because ideational tasks benefit from exploration of varied and unfamiliar terrains in the conceptual space of ideas (Pelled et al. 1999), whereas coordination tasks require team members to be on the same page about who does what and when (Converse et al. 1993). Thus, discursive diversity during ideational tasks will equip group members with new ways of interpreting the shared problem and enable them to recombine ideas in ways that yield novel solutions. Conversely, discursive diversity during coordination tasks will sow confusion and make it harder for group members to find the common ground needed for smooth implementation. Thus, we anticipate:

HYPOTHESIS 1: *The capacity to modulate discursive diversity such that high levels are exhibited during ideation and low levels during coordination will be positively related to team success.*

Congruent Modulation, Discursive Diversity, and Team Performance

Having argued that high-performing teams are ones that modulate group discourse in ways that reinforce shared meanings when they are coordinating and encourage the expression of diverse meanings when they are ideating, we turn next to considering the *process* by which this collective adjustment unfolds. In particular, we consider how the degree of congruence in group members’ adjustments of discursive diversity relates to team performance.

Given prior research on behavioral synchronization in groups, we anticipate that teams are more likely to perform well when they are “in sync” with one another in adjusting their expressed meanings to meet shifting task requirements. Moreover, we propose that the degree to which members are aligned in their adjustments matters for team success independent of the team’s level of discursive diversity for a given task type. Consider two possible pathways for a team to make the same amount of adjustment in discursive diversity to a change in task requirements. In both scenarios, assume that the team’s overall discursive diversity is the same across task types. In one scenario, which we label *incongruent modulation*, a small number of group members make dramatic shifts in the meanings they express, while the majority of the group makes little to no adjustment. In incongruent modulation, the group-level variance in members’ within-person semantic shifts is relatively high. In the second scenario, which we call *congruent modulation*, all team members make shifts in the meanings they express that are of comparable magnitude. Congruent modulation is instead characterized by relatively low group-level variance in members’ within-person semantic shifts.

Building on social entrainment theory (McGrath and Kelly 1986, Ancona and Chong 1996, 1999), we argue that congruent modulation is more likely to be associated with team success than

is incongruent modulation. Support for this view comes from prior work on social entrainment, which suggests that behavioral synchronization between group members can arise not only through shared attention to an external “pacer,” such as an impending deadline, but also endogeneously through an interactional rhythm that emerges between group members (e.g., Ancona and Chong 1996, 1999, Chapple 1974, McGrath et al. 1984).

Studies of entrainment in groups have focused on the performance benefits that arise when team members’ activities are coordinated. For example, time constraints such as project deadlines alter the rhythms of work group activity in ways that influence team performance (Kelly et al. 1990, Kelly and Karau 1993, Kelly and McGrath 1985). Similarly, orderly turn-taking in group discussion and problem solving is associated with positive group outcomes (e.g., Warner 1979, 1992). And, across contexts as diverse as music and sports, high-performing teams are ones whose members are closely aligned with one another as they increase or decrease their physical effort (McGrath et al. 1984).

Beyond group members’ activity levels, social entrainment can also occur in the meanings that group members express through interaction. In a study of engineering design teams, Brereton (1996) showed that group discussions proceed through alternating cycles of focused conversation about specific issues and transitional episodes where teams switched between topics. Similarly, Reid and Reed (2000) showed that design teams become entrained to alternating rhythmically between discussions that are primarily conceptual and ones that are primarily figurative. In both studies, group members were coordinated with each other in changing their expressed meanings.

Integrating the idea that groups can become entrained in shifting from states of semantic convergence to divergence (and vice versa), as well as the insight that such alignment is generally beneficial for team performance (e.g., Warner 1992, Chapple 1974, Cannon-Bowers et al. 1993), we suggest that high-performing teams will not only match their levels of discursive diversity to shifting task requirements but will also make individual-level modulations that are collectively congruent. Stated as the inverse, we expect:

HYPOTHESIS 2: *Incongruent modulation—that is, high group-level variance in the within-person semantic shifts made by team members to changing task requirements—will be negatively related to team success.*

Method

Research Setting and Data

Our research setting is Gigster (gigster.com), an online platform on which freelance software developers produce on-demand software for individual and corporate clients. Unlike many two-sided platforms that match individual freelancers to clients who need help on focused, independent projects,

this platform assembles individual freelance developers into temporary teams, headed by a team leader, and assigns them to longer-term projects that require complex, interdependent work. The freelancers on this platform are distributed around the globe and work on a variety of projects ranging from mobile to web application development. The projects are generally knowledge-intensive, requiring high levels of creativity, technical problem-solving, and interpersonal coordination. Software projects on this platform are significant in scope and vary in cost from tens to hundreds of thousands of dollars (and upwards of one million dollars at the extreme).

Our dataset comprises 117 teams, representing 421 unique individuals (36% female), and spans the period from early 2015 until late 2017. A typical team had 5 members and consisted of one project manager, at least one backend, frontend or “fullstack” engineer, a designer, and a user interface expert. Depending on the type of project, teams sometimes also included writers, natural language processing engineers, and other types of specialized professionals. Among teams in our data, projects lasted 159 days on average (median: 150 days) and were structured in milestone phases that lasted two weeks, on average (mean: 14 days; median: 14 days). To join the platform, professionals had to pass a variety of technical interviews designed to verify their expertise. On average, the members of an individual team represented 3.6 countries (median: 3). 42% of individuals in our sample listed their country of origin as located in North America. Another 13% hailed from Asia, followed by 12% from Europe. The remaining 23% resided in Latin America, Africa, and other parts of the world.

Because they were geographically distributed and lacked any physical office space, team members communicated almost exclusively via an online instant messaging tool called Slack. We had access to the entire set of teams’ Slack archives—over 800,000 messages. Each message was timestamped and attributable (via anonymized identifiers) to its author. On average, teams exchanged 1,873 Slack messages in public channels throughout their lifespan (median: 1,220). We conducted informal interviews with Gigster’s senior leaders and team leaders, who uniformly reported that team communication happened nearly exclusively through Slack. One senior leader described the reasons for this: “Almost all team conversation happens on Slack. It’s a useful tool because we run global teams, and Slack allows for real-time and asynchronous communication within one platform. It also allows for easy file-sharing on projects.” Multiple informants highlighted that team members consistently relied on Slack as opposed to other tools because of the importance of having “everything in one place” to facilitate team collaboration. Informants also said that team members had an incentive to use Slack because it provided a transparent archive of team processes and events that was used to triage some rare cases of disputes.

In addition to Slack messages, we had access to data on team member characteristics—functional role, gender, and country of origin—as well as overall team performance in meeting its various

project milestones. Together, these data constitute a rich and continuous history of teams’ internal dynamics and outcomes.

Dependent Variable

The timely delivery of milestones is the most critical performance measure for teams on this platform. Company executives explained that clients prioritize timeliness, and this is the key metric used to evaluate individual freelancers. The final project deliverable, as well as the deliverables for each milestone and dates for corresponding deadlines, are agreed upon between the project manager and the client before the project starts. Timely delivery signals both effective coordination among team members, as well as high output quality, since a client must approve or reject the team’s deliverables at each milestone deadline. If the agreed-upon deliverables for a given milestone are deemed by the client to be of poor quality or incomplete, that milestone is marked within the company’s system as delayed. Teams are allowed to proceed from one milestone phase to the next only after the client approves a given milestone’s deliverables as satisfactory. Team members are paid a pre-agreed sum upon the approval of each milestone, as well as upon successful completion of the project. Members of teams that do not deliver on time may experience financial penalties or limited opportunities to join lucrative projects in the future.³

Independent Variables

Discursive Diversity. We developed our language-based measure of discursive diversity using word embedding models, a neural network-based family of unsupervised machine learning methods for representing words in a high-dimensional vector space. Previous work demonstrates that word embedding models are particularly useful for capturing semantic relationships between words. These relationships correspond to the underlying categories of meaning that inform speakers’ language use. Garg et al. (2018), for example, demonstrate that different occupations’ semantic gender associations, as inferred from word embedding models applied to English books published throughout the twentieth century, correspond to these occupations’ historical gender compositions. Similarly, Kozlowski et al. (2019) illustrate how different lifestyle activities are associated with class, race, and gender identities. Thus, word embeddings offer holistic and meaningful insights into numerous dimensions of meaning contained in language that prior methods have been unable to capture.

We began by fitting a word embedding model to the entire corpus of teams’ Slack archives (vocabulary size 9,500). We pre-processed the Slack data according to standard procedures in

³We also had access to client satisfaction scores for teams; however, after consultation with the leaders of the platform, we opted not to use this measure as a performance outcome because there is little variance in satisfaction among the clients that choose to report these scores. As we understand it, timeliness of milestone completion is the most economically consequential outcome for teams on this platform.

natural language processing and trained our embedding model using a Python implementation of Word2Vec, a popular implementation of CBOW word embedding models.

Following standard practice, we trained a 200-dimensional model using a window size of 10 words (Mikolov et al. 2013). From this training process, we obtained one 200-dimensional vector of coordinates for each word in the corpus that represented the word’s location in semantic space. As described previously, the dimensions of an embedding space represent latent features underlying language use in the training corpus. While the dimensions do not possess qualitatively interpretable meanings in and of themselves, the dimensions are informative in the sense that words with more similar meanings are located closer to each other.

Our discursive diversity measure can be applied to time windows of varying lengths. While it provides significantly greater temporal granularity than self-report measures do, at too refined a time resolution the volume of language per individual is too sparse to generate meaningful embedding centers of mass. Because communication on Slack tends to be brief and conversational, with individual posts often comprising just a few words, we applied the measure at the daily level and then computed mean discursive diversity across project stage intervals as described below. In other words, we computed an average daily discursive diversity score for each of the three task stages in a milestone.

We now turn to describing how the discursive diversity measure was computed for each team and for a given time window. Let I be a team of N individuals, and W_{it} denote the set of words expressed by individual i during time period t . We define $\overline{W}_{it} = \frac{1}{|W_{it}|} \sum_w v_w$ as the embedding centroid for individual i during period t , where v_w is the embedding vector representation for word w . $\overline{W}_{it}$ represents i ’s embedding center of mass during time period t . This is the individual’s mean position on each dimension of the embedding space as derived from her use of language during that time.

We define the embedding distance between two individuals, i and j , during time t , as the cosine distance between their respective embedding centers of mass:

$$d(W_{it}, W_{jt}) = 1 - \cos(\overline{W}_{it}, \overline{W}_{jt}) \quad (1)$$

where $\cos(A, B) = \frac{AB}{\|A\|\|B\|}$. Using this distance metric, we define the group’s overall discursive diversity as the average pairwise embedding distance between all members of the group:

$$DD_t = \frac{1}{N^2} \sum_{i \in I} \sum_{j \in I} d(\overline{W}_{it}, \overline{W}_{jt}) \quad (2)$$

Figure 1 below illustrates how we computed a discursive diversity score based on teams’ Slack conversations. We first computed one vector to represent each chat message by computing the

centroid of the individual vectors associated with the words in the message (Panel A, Figure 1). Panel B of Figure 1 shows how we applied this approach to arrive at a team discursive diversity score: First, we computed one centroid representing each team member’s messages on a given day. Then, we computed the pairwise cosine distances between these person-vectors. Lastly, we computed team discursive diversity as the average of these pairwise distances.

[FIGURE 1 ABOUT HERE]

Figure 2 below shows a stylized example of how discursive diversity scores are computed for a team at two points in time, one at which the team are relatively less discursively diverse and one at which the team are more discursively diverse.⁴ At time one, shown on the left-hand side of Figure 2, discursive diversity is low, as all three members use semantically similar, and in some cases even identical, words (bolded). As a result, the three centroids representing these individuals’ utterances are located close to each other in embedding space. Since the pairwise cosine distances between the person-centroids are relatively small, the team’s discursive diversity score at time one is relatively low. In contrast, the right-hand side of Figure 2 shows the same team at a time when discursive diversity is higher: The words that team members utter are substantively different in their meanings, resulting in person-centroids that are relatively farther apart from each other in embedding space. Consequently, the cosine distances between the person-centroids and the team’s discursive diversity score are higher.

[FIGURE 2 ABOUT HERE]

Discursive Diversity by Task Type (i.e., Coordination versus Ideation). Given that our theory focuses on the effects of discursive diversity on different types of tasks the team performs, we turn next to describing how teams on this platform structure their work. All teams in our data operated according to the “Scrum” framework, which is a common project management approach in the software and technology industries (Schwaber 1997). Scrum allows teams to break their project work down into “mini projects” that can be completed within repeated, time-delimited iterations—so-called “sprints.” Sprints typically last between two weeks and one month. Each sprint culminates in a deliverable or “milestone,” which represents an incremental piece of progress toward the project’s final output.

Sprints typically progress through three stages of roughly equal length, each of which involves a distinct set of tasks: sprint planning, daily Scrum work, and sprint review and integration. During sprint planning, the team creates a shared understanding of the milestone goals, the actions necessary to accomplish these goals, and the high-level tasks and responsibilities that will be

⁴For ease of visual presentation, the graphs in Figure 2 represent the 200-dimensional embedding space as a two-dimensional space of two principal components, PC1 and PC2.

assigned to team members. Overall, the planning stage focuses more on coordination than on ideation.

In the second stage, the team identifies how best to accomplish the goals established in the planning stage and troubleshoots new and unexpected challenges that arise as the work unfolds. Team members collectively brainstorm solutions to technical problems, ask questions, and provide feedback on each other’s ideas and solutions. Although daily Scrums involve a mix of ideation and coordination, success in this stage is more about the former—that is, creative problem solving and identifying new or unanticipated ways to achieve milestone goals.

The last stage of a sprint consists of review and integration, when team members come together to integrate their individual outputs, review what has been accomplished, and discuss how to complete outstanding tasks. Overall, this last stage involves more coordination than ideation, given that teams have to close off new ideas and instead get aligned to deliver on milestone goals. To recap: in this setting, sprint planning (stage 1) and sprint review and integration (stage 3) primarily involve coordination tasks, whereas daily scrum work (stage 2) is more focused on ideation tasks. Figure 3 provides a visual representation of the project timeline and milestone stages for a typical team.

[FIGURE 3 ABOUT HERE]

Given that our interviews of project managers and executives on the platform suggested that the three sprint stages were of roughly equal length, we used milestone-thirds as a proxy for the task that the team was engaged in at a given point in time. That is, we assume that the first third of any milestone corresponds to sprint planning, the second third to daily Scrum work, and the final third to review and planning.

This assumption is informed by the informal interviews we conducted with Gigster’s leadership team and project team leaders about the task structure of milestones. Our interviewees uniformly reported that teams consistently followed the Scrum methodology, moving from coordination to ideation and then back to coordination in each milestone. As one senior leader described, “The milestones on a project look really similar in terms of structure. Milestones start out with a team alignment phase or planning phase for what’s going to be covered within the milestone. We try to allocate the middle section of a milestone to as much problem-solving and creative thinking as possible. And then the key is that the team comes back together at the end of the milestone to review what’s been done, tie everything together, and plan for the next milestone.” One designer, who later worked as a team leader, said: “Each milestone is like its own mini project, with its own deliverables that need to get done. During the first part of the milestone, the leader and their team plan what needs to get done and who will do what. Then, in the middle phase, everybody goes and does those things. This entails a lot of joint problem-solving, coming up with good ideas, and

helping each other troubleshoot. For example, people will critique designs collectively to make them better or debug code together. In the last phase, everybody comes back to put all the puzzle pieces together ahead of delivery. We also use that time at the end to run what’s called a retrospective, looking back at what got done, whether it went well, and what remains to be done going forward. Each of these three stages takes about the same amount of time.”

Beyond this qualitative evidence, to further validate that coordination activities were indeed more prevalent during the first and last thirds of a milestone than in the ideation-focused middle third, we conducted a t-test on the relative prevalence of words that are closely associated with coordination activities. We selected the following coordination-related words based on our knowledge from qualitative analyses of teams’ Slack data: *timeframe*, *prioritizing*, *budget*, *shipping*, *deliverable*, *approval*, *progress*, *planning*, *reporting*, *communicating*, *handoff*, *reviewing*, *approving*, *scope*, *timeline*, *kickoff*, *handover*, *creep*, *todo*, *trello*, *card*, *asana*.⁵ Our t-test compared the relative prevalence of these words in the coordination stages (stages one and three) with their relative prevalence in the ideation stage (stage 2). Coordination words were significantly more prevalent in coordination stages than in ideation stages ($t = 3.499, p = 0.0005$).

Degree of Congruence in Individual-level Modulations. Hypothesis 2 focuses on the extent to which teams are congruent in their modulations of discursive diversity to shifting task requirements. To test this hypothesis, we first developed a measure of within-person discursive range, which captures the semantic distance between an individual’s language use across the three milestone stages. In particular, we computed for each individual in each milestone the average distance between the centroids representing her language use during stages 1 and 2 and during stages 2 and 3, respectively. Given that Hypothesis 2 focuses on the collective *incongruence* of individual-level modulations, we defined *incongruent modulation* as the team-level variance in members’ within-person discursive range from coordination (stage 1) to ideation (stage 2) and from ideation (stage 2) to coordination (stage 3).

Analytical Strategy

We estimated conditional logit models of whether or not a team achieved timely delivery of a given milestone (which we denote as “success”). All models described in this section include team fixed effects to account for unobserved heterogeneity in project complexity and team characteristics such as skills, ascriptive characteristics, past work experiences, and personality traits that might account for variation in team performance. Given that team leaders rarely turned over during a given project and, if they did, only left in between milestones, our focus on the milestone as the unit of analysis

⁵Our analyses were conducted using stemmed versions of these words and of the Slack data to ensure that all variants of the target words were included and detected.

and the inclusion of team fixed effects also address concerns about unobserved heterogeneity in team leaders’ past experiences and team members’ past collaboration histories. All models also include a control for milestone number to account for differences in the likelihood of success for earlier versus later milestones and milestone length fixed effects to account for unobserved differences in the likelihood of success for shorter versus longer milestones.

Despite the inclusion of team and milestone length fixed effects and a control for milestone number, we cannot, however, fully rule out the threat of endogeneity—a point to which we return in the discussion section. We clustered errors at the team level to account for the non-independence of milestone-level observations for a given team.

Results

Discursive Diversity: Validating the Word Embedding Model

As a first step, we sought to validate that our word embedding model did indeed uncover and appropriately represent the meanings expressed in team members’ Slack communications. There are two common approaches to evaluating the validity of word embedding models: most-similar queries and word analogy tasks. In most-similar queries, the model is asked to return the words that it learned to be most similar to the vector of a given target word. For example, in software development, the word “bug” usually refers to a programming issue, whereas it is more likely to refer to an insect in other contexts. Our model evaluated the most similar words to “bug” to be “issue,” “crash,” and “problem,” demonstrating that the meaning of “bug” was accurately captured in context-relevant way. Similarly, the most similar words to “sweet” were “intense,” “dope,” and “yay,” while the most similar words to “dude” were “man,” “bro,” and “yessir.” These examples demonstrate an important advantage of custom-trained word embeddings over non-customized approaches to building language models: They capture not only semantic relationships but also relationships between somewhat idiosyncratic cultural schemata that are used in a given context—in this case, freelance software development teams. We conducted a wide range of most-similar queries for target words from within and outside the software development context and found that the model appeared to capture their meanings in contextually appropriate ways.

Second, Mikolov et al. (2013) show how mathematical operations in the vector space produced by an embedding model can be used to solve analogical reasoning problems and can serve as a further check of model validity. For example, the authors use their model to evaluate the question, “Germany is to Berlin as France is to ____?”. The answer is given by $v(Berlin) - v(Germany) = x - v(France)$, or $x = v(Berlin) - v(Germany) + v(France)$. Provided that the model correctly learned vector representations of words, the answer (in this case, $x = Paris$) that the model returns will be given by the word vector that is closest to the coordinates that are obtained from this equation.

Applying this approach to our word embedding model provided further evidence that our model performed well at capturing not just common semantic relationships but also certain meanings that are idiosyncratic to the context of freelance software development. We tested our model through a number of word analogy tasks, some examples of which are shown in Table 1. We interpret the results of both types of validity checks, most-similar queries and word analogy tasks, as evidence that our model appropriately captured context-relevant semantic relationships between words.

[TABLE 1 ABOUT HERE]

Descriptive Statistics

Teams varied with respect to average daily discursive diversity during a milestone phase (*mean* : 0.645, *SD* : 0.123) and with respect to average within-person discursive range during a milestone phase (*mean* : 0.86, *SD* : 0.436). Team discursive diversity and average within-person discursive range were significantly positively correlated ($r = 0.24, p < 0.001$). This makes intuitive sense, as we would expect team members who are more semantically flexible to show greater levels of semantic distance from one another, on average. Discursive diversity was weakly correlated with team gender diversity ($r = 0.09, p < 0.05$) but not with team diversity with respect to functional roles or members’ countries of origin.⁶ The average milestone phase lasted 14 days (*SD* : 9.02). On average, teams had 5 members, of whom 3.45 participated in the Slack conversation during a given milestone (*SD* : 1.6). Teams were relatively diverse with respect to functional roles and members’ countries of origin, and relatively less diverse with respect to gender. Summary statistics and bivariate correlations for the variables in our main models are reported in Table 2.

[TABLE 2 ABOUT HERE]

Variation in Discursive Diversity across Teams’ Life Cycles and Milestones

To assess variation in discursive diversity, we first examined the degree to which it varies across teams’ life cycles—that is, from the beginning of the project, when team members first convene to begin working together, to the very end of the project when the team disbands. Figure 4, Panel A, plots team discursive diversity as a function of team life stage, where life stages of 0 and 1 correspond to the start and end points of the team’s life, respectively. Each of the grey lines in the figure represents one team from a sample of 20 randomly selected teams. Since teams complete multiple milestones throughout their life cycles, Figure 4, Panel B, plots team discursive diversity as a function of milestone life stage for 20 randomly selected milestone phases, where life stage 0

⁶We measure three types of team demographic diversity: Functional role, gender, and country of origin diversity. Each was quantified using a standardized Blau index (also referred to as an inverse Simpson or Herfindahl index), where values closer to 1 indicate greater heterogeneity and values closer to 0 indicate greater homogeneity with respect to the focal characteristic (Gibbs and Martin 1962).

corresponds to the start of a milestone phase and stage 1 corresponds to the milestone due date. The blue lines represent the mean level of discursive diversity at a given team life stage (Panel A) and the mean level of discursive diversity at a given milestone life stage (Panel B), respectively.

[FIGURE 4 ABOUT HERE]

Figure 4 illustrates the core advantage of our approach to measuring cognitive diversity relative to traditional survey-based measures: Even if collected at a few points in time during a team’s life cycle, self-reports of cognitive diversity would simply be unable to capture the fine-grained temporal variation that language-based measures can uncover.

Figure 5 shows the distribution of teams’ mean levels of daily discursive diversity. As the plot indicates, discursive diversity varied substantially not only within, but also across, teams.

[FIGURE 5 ABOUT HERE]

Main Results

Table 3 reports our main results. To test Hypothesis 1, we began by investigating whether teams’ mean discursive diversity *across the milestone as a whole* is predictive of performance. Model 1 shows that the relationship between mean diversity levels and performance is not significant.

[TABLE 3 ABOUT HERE]

The picture changes, however, once the type of task being undertaken by the team is taken into account. Our first hypothesis suggests that ideational discursive diversity (during stage 2) will be positively associated with success, while coordination discursive diversity (during stages 1 and 3) will be negatively associated with success. Models 2, 3, and 4 predict performance based on discursive diversity in each of the three milestone stages individually. The results in Model 5, our main model that considers the effects of discursive diversity in all three stages, are consistent with this expectation.⁷

These effects are substantively meaningful, as calculations of teams’ predicted probabilities of success based on Table 3, Model 5, show. In stages 1 and 3 (i.e., the coordination stages), teams with discursive diversity levels one standard deviation above the mean have a 9 percent lower predicted probability of success compared to teams with discursive diversity levels one standard deviation below the mean (*probability* = 0.46 vs. *probability* = 0.37). The story reverses, however, in Stage 2, the ideation stage: teams whose discursive levels are one standard deviation above the mean have a 10 percent higher predicted probability of success relative to ones that are one standard deviation below the mean (*probability* = 0.47 vs. *probability* = 0.37). Together, these results provide strong support for Hypothesis 1.

⁷In (unreported) robustness checks, we found the same basic pattern, where discursive diversity was negatively related to performance in the early and late stages, but positively related to performance in the middle stages of a milestone phase, for different splits of milestone phases—for example, when time was segmented into quarters or fifths.

Hypothesis 2 anticipates that teams exhibiting incongruent modulation—that is, ones characterized by high group-level variance in members’ within-person discursive range—are less likely to achieve success. In line with this expectation, Model 6 in Table 3 shows that the total variance in team members’ within-person semantic shifts from stage 1 (coordination) to stage 2 (ideation) and from stage 2 to stage 3 (coordination) is negatively related to team success. This relationship holds when the stage-wise discursive diversity variables are included in the fully saturated Model 7. Again, the predicted probabilities derived from 3, Model 7, demonstrate that this effect is meaningful: teams that are one standard deviation below the mean in their variance in within-person shifts between stages have a 13 percent higher predicted probability of success compared to teams that are one standard deviation above the mean (*probability* = 0.53 vs. *probability* = 0.40). Overall, these results lend support for Hypothesis 2.

Extensions

Having demonstrated that successful teams are ones that modulate discursive diversity to match changing task requirements and that exhibit congruent modulation in making these adjustments, we turn next to exploring the factors that underlie this collective capability. In particular, we conducted two supplemental analyses—one focused on the role of the leader and the other on how individual group members’ semantic range relates to the group-level discursive diversity.

Unpacking the Role of Team Leaders in Modulating Discursive Diversity. In our setting, team leaders serve as the main point of contact between the client and team members. They are responsible for ensuring that the team attends to the right issues at the right time, that deliverables are met in a timely fashion, and that the team is communicating and collaborating effectively. Given the leader’s central role, we conducted a post hoc analysis that provides suggestive evidence of the leader’s importance in shaping the relationship between discursive diversity and team performance. In particular, we decomposed discursive diversity into the diversity that exists between the leader and followers and diversity that exists between only followers. The results are shown in Table 4. Model 1 and Model 2 indicate that only the former form of discursive diversity—that between leaders and followers—appears to matter for performance. This finding is consistent with the notion that successful team leaders create a context in which followers hew closely to the leader during coordination stages and feel free to deviate from the leader during ideation stages.

Interestingly, as Model 3 in Table 4 shows, leader-to-follower discursive diversity in the ideation stage is not significant once between-follower discursive diversity in that stage is controlled for. A possible interpretation of this result is that effective leaders downplay the hierarchical distance between themselves and team members during ideational discourse. During discourse about coordination, however, they maintain hierarchical distance with subordinates. While our data do not

allow us to pin down the mechanisms underlying these findings, we hope that they will prompt further work on the role of leaders in teams’ capacity to modulate their levels of discursive diversity.

[TABLE 4 ABOUT HERE]

Antecedents of Discursive Diversity. As mentioned previously, our results show that demographic diversity among team members is not a meaningful predictor of discursive diversity in our data. Only gender diversity was correlated—though weakly—with discursive diversity. To better understand the antecedents of team discursive diversity, we investigated the relationship between individuals’ within-person discursive range—the measure described above of individuals’ semantic flexibility between task types—and discursive diversity. We anticipated that teams whose members exhibit greater within-person discursive range, on average, would also show higher team discursive diversity levels. As reported in Table 4, Model 5, this is indeed what we found. In other words, teams with members who exhibit greater semantic flexibility achieve higher mean levels of discursive diversity across all stages of a milestone. Yet, as Model 4 illustrates, semantic flexibility by itself does not support team performance. Juxtaposing this result with those in Table 3, we conclude that individual-level semantic flexibility equips teams with the capacity to exhibit discursive diversity; however, only the teams that have the capacity to modulate their levels of discursive diversity to match shifting task requirements are able to translate this capacity into successful outcomes.

Discussion

The goal of this article has been to bring conceptual clarity and new empirical evidence to bear on a longstanding question in organizational theory: How does cognitive diversity among members of a group influence their performance? Prior work has assumed that group cognitive diversity confers the benefits of creativity and innovation but, at the same time, imposes the costs of misalignment and strained coordination (Milliken and Martins 1996). Thus, for a given set of members, the prevailing assumption has been that groups can shine at creativity or at implementation but generally not at both.

Building on the insight that meaning is produced collectively and dynamically through interaction (Berger and Luckmann 1967) and that interactions between group members enable them to adjust their understandings of shared problems and potential solutions (Thompson and Fine 1999), we developed a theoretical account of how groups can navigate the performance tradeoff of group cognitive diversity. In doing so, we introduced the constructs of *discursive diversity*—the divergence in meanings conveyed by group members in a given set of interactions—and *congruent modulation*—low group-level variation in the degree to which individual members modulate their language as tasks change. We used a deep-learning linguistic method (Mikolov et al. 2013) and data

from 117 software development teams on an online platform to develop novel measures of these constructs. Unlike prevailing survey-based methods for assessing group cognition, our method makes it possible to measure fine-grained temporal fluctuations in group members’ cognitive diversity as they face changing task requirements, as well as the process by which they collectively adjust group cognition.

We theorized and found support for the notion that high-performing teams are ones that have the collective capacity to modulate their levels of discursive diversity to match their changing task requirements: they exhibit higher discursive diversity when engaged in ideational tasks and lower discursive diversity when performing coordination tasks. We further found that teams exhibiting congruent modulation in making these adjustments—that is, those with lower group-level variance in members’ within-person semantic shifts—are more likely to achieve success.

We turn next to considering the boundary conditions of our theory and the extent to which it might generalize to other contexts. Because our arguments focus on a team’s capacity to modulate its levels of discursive diversity to shifting task requirements and the degree to which these adjustments are congruent across team members, our theory would generally not apply to teams that experience no variation in task requirements. For example, we anticipate that a team that is engaged exclusively in coordination—for example, repetitive mechanical assembly tasks—will achieve success by maintaining low levels of discursive diversity throughout its lifecycle. Conversely, a team that focuses on pure ideation—for example, a screenwriting team for a comedy sketch series that develops a loosely connected storyline but then hands off implementation of the sketch (e.g., choosing wardrobes and sets) to others—will be successful insofar as it constantly operates at high levels of discursive diversity. Yet most teams in organizational settings do not fit neatly into these ideal types; rather, the typical team has to switch at varying points between ideation and coordination. Our theory can be readily extended to such teams, even when they do not fit the pattern of oscillation between coordination and ideation observed in our empirical setting. Consider, for example, creative teams such as those in the advertising industry. Our theory would predict that, to achieve success, such teams need to exhibit high levels of discursive diversity early in their lifecycle while they are brainstorming new ideas but low levels toward the end of their lifecycle as they shift from idea generation to idea implementation.

Shared Cognition in Groups

Our first core contribution is in introducing the construct of discursive diversity, a manifestation of group cognitive diversity, which departs from prevailing constructs in three main ways: (1) it draws inferences about group members’ ideas and meanings from the language they express in routine interaction rather than probing those meanings through abstract survey questions; (2) it

does not make a priori assumptions about which dimensions of meaning are relevant in a given group interaction and potentially consequential for success; and (3) it considers how the full set of meanings that group members convey to one another—rather than a privileged set of categories (e.g., mental models about team tasks, processes, or properties of the team itself)—can contribute to convergence or divergence in a group’s shared cognition.

In developing and operationalizing this construct, we challenge the prevailing view in research on groups and teams of diversity as a “double-edged sword” that necessarily aids creativity at the expense of coordination (Milliken and Martins 1996). By bringing in the roles of group interactions that give rise to shared cognition and the capacity of groups to modulate their discourse, we demonstrate that groups can negotiate the performance tradeoff of group cognitive diversity by developing the collective capacity to modulate their levels of discursive diversity to match their changing task requirements.

We further contribute to research on diversity in teams by demonstrating that our measure of cognitive diversity is not strongly correlated with demographic variables—with the exception of a weak but significant correlation with gender diversity—yet still matters for team success. Although we acknowledge that there may be broader organizational and societal benefits to assembling demographically diverse teams, this suggests that demographic diversity by itself may be insufficient for generating the breadth of ideas needed for team success. It also points to the possibility of using discursive diversity as a basis for team construction. For example, a professional service firm that is assembling a client service team could analyze the communication patterns of prospective team members on prior projects and select individuals in part based on the discursive range they previously exhibited and their demonstrated capacity for modulation and for making congruent shifts.

Moreover, we contribute to the literature on social entrainment by introducing a novel measure of the extent to which group members are “in sync” with one another in making semantic adjustments to changing task demands. Most of the prior work on social entrainment in groups has primarily focused on the synchronization of group members’ activity levels, for example with respect to physical effort (e.g., McGrath et al. 1984). Only a few studies have examined cognitive entrainment with respect to cyclical patterns in the substantive focus of group discussions (e.g., Reid and Reed 2000, Brereton 1996), but these studies did not examine the extent to which individual group members were aligned with one another when shifting the group’s overall discourse. We show that congruent modulation is associated with higher performance, suggesting that the *process* by which groups modulate their levels of discursive diversity can also matter for group success.

Finally, we conjecture that discursive diversity may play an important role in the emergence of other facets of group cognition. A leading candidate is group transactive memory—a repository

that emerges within a team to encode, organize, and share knowledge from group members' different domains of expertise (Wegner 1987, Ren and Argote 2011, Reagans et al. 2016, Aggarwal and Woolley 2019). Transactive memory systems are more likely to emerge, and are more effective, when individuals engage in repeated interactions and form close relationships (Wegner et al. 1991). We believe that a promising path for future research is in investigating how discursive diversity influences this process. High levels of discursive diversity may, for example, enable teams to more efficiently understand and map the full range of knowledge, expertise, and skills that exist among members. Yet it may also make it harder for teams to successfully encode and retrieve this knowledge. It may even be possible to develop language-based measures of a group's transactive memory system so that its interrelationship with discursive diversity can be directly examined.

Collective Meaning Making in Groups

Next, we believe that the methodological approach we develop to measure discursive diversity can be extended beyond the context of goal-directed teams to understand collective meaning making in groups more generally—from children negotiating social roles on the playground to analysts making sense of turbulent financial markets (Knorr Cetina and Bruegger 2002). For example, we suspect that discursive diversity may be central to what Collins (2014) refers to as an interaction ritual chain—group interactions that involve the mutual focus of attention, as well as *emotional* entrainment in the form of bodily synchronization and mutual arousal. Such interactions imbue group members with emotional energy and strengthen group attachment.

We suspect that group members who identify strongly with a group will generally exhibit low levels of discursive diversity, reflecting their shared understandings of the world. However, when they are engaged in interaction ritual chains that are characterized by high levels of discursive diversity, we anticipate that group members will experience heightened emotional entrainment and thus feel even more strongly identified with the group. Moreover, although Collins initially described this process unfolding in face-to-face interactions such as spectators at sporting events or employees taking smoking breaks at work, more recent work shows that interaction ritual chains can also unfold in online settings such as employees of a multinational company discussing its core values and beliefs (DiMaggio et al. 2018). Thus, it may be possible to apply our approach to measuring discursive diversity to corpora of interactional language use among various types of groups and forecast when interactions will strengthen versus weaken individuals' identification with and attachment to the group. Applications of such an approach range from predicting when political polarization will intensify to forecasting when organizational faultlines are likely to emerge.

Teams and Organizational Ambidexterity

Finally, the construct of discursive diversity can shed new light on the role that teams play in helping the organizations to which they belong navigate the tensions of exploration and exploitation (March 1991). In particular, a prominent line of research has examined the antecedents of organizational ambidexterity—the ability to simultaneously pursue both incremental and disruptive innovation (Tushman and O’Reilly 1996). Prior work has highlighted three pathways by which organizations can achieve ambidexterity: sequential, or oscillating back and forth between periods of exploration and exploitation; structural, or pursuing both objectives simultaneously by separating the two sets of activities and the capabilities needed to execute them into distinct organizational subunits; and contextual, or creating a context in which individuals can exercise appropriate judgment about when to pursue exploration versus exploitation (O’Reilly and Tushman 2013, Rogan and Mors 2014).

Research on the role of teams in fostering organizational ambidexterity has tended to focus on the structural pathway and highlighted the role of top management teams in balancing the competing interests of the units dedicated to building new capabilities versus those focused on harvesting existing ones (Jansen et al. 2008, Carmeli and Halevi 2009, Mihalache et al. 2014). With the introduction of discursive diversity, we open an alternative route to organizational ambidexterity: equipping teams throughout the organization—not just at the top—with the skills needed to match their levels of discursive diversity to their task requirements. In a sense, organizations possessing such teams can be thought of as integrating the sequential—oscillating between high and low levels of discursive diversity—and contextual—matching those oscillations to cycles of exploration versus exploitation—pathways to ambidexterity.

Limitations and Future Directions

Given the data available to us, this study has certain limitations, which point to useful avenues for future research. First, although we relied on qualitative interviews with project managers and platform leaders and qualitatively coded transcripts of Slack communications, we ultimately had to rely on time-based proxies for ideational versus coordination tasks. Whereas the teams in our data uniformly followed the Scrum framework, in other settings teams vary in how they structure their work. To generalize our approach to other types of teams, one would ideally pair electronic communications data with other artifacts that reveal the nature of tasks the team is engaged in at different points in time. In the context of software development, for example, researchers could potentially tap into teams’ Github repositories or project management software tools to inspect how the nature of tasks shifts over a team’s lifecycle.

Although our outcome measure—timely delivery of milestone objectives—was economically consequential in our setting, we acknowledge that it is nevertheless a crude indicator of team success.

Future research would benefit from pairing objective indicators of success (e.g., timeliness, productivity, and profitability) with subjective measures of team learning, relationship quality, and well-being to build a fuller account of the outcomes that are shaped by discursive diversity.

Despite our use of models that include team and milestone length fixed effects and thus account for unobserved, time-invariant heterogeneity between teams and types of milestones, we are not able to make strong causal claims with our empirical setup. For example, it is conceivable that teams that are confident they will achieve milestone success might feel more free to communicate in discursively diverse ways, whereas teams that know they will miss a milestone might narrow their ambitions and communicate with less discursive diversity. Given that any potential shifts in discursive diversity as response to anticipated performance are most likely to occur in later milestones and during longer milestones, the inclusion of the control variable for milestone number and of milestone length fixed effects in our models partially addresses this concern. To fully overcome the threat of reverse causality, we suspect that it will be necessary to design interventions that nudge teams to communicate in more or less discursively diverse ways, randomly assign teams to receive these different treatments, and then compare the performance of the two sets of teams (Hauser et al. 2018).

Another promising avenue for future research lies in the question of whether discursive diversity would have similar performance effects in settings where teams meet in person. This question is pertinent because, despite the growing prevalence of remote collaboration practices and tools across industries, significant portions of work in many organizations are still accomplished in face-to-face interactions. Nonverbal behavior and paralinguistic cues, such as facial expressions, body language and tone of voice, can convey important cultural information in such interactions (ten Brinke et al. 2016, Schroeder et al. 2017). Data sets that encompass multiple modes of communication are needed to enable scholars to study how the diversity of nonverbal and paralinguistic expressions might work alongside textual discursive diversity to influence team performance.

Finally, our findings raise a number of intriguing questions about how teams can build the capacity for modulating their levels of discursive diversity. To what extent are teams aware of impending shifts in task requirements? How do they achieve a state of entrainment in making these adjustments? What practices of leaders and team members enable the team as a whole to dial discursive diversity up or down as needed?

Conclusion

This study illuminates how shared cognition in groups arises through interaction and manifests in the form of discourse. It introduces a novel language-based method to measure fine-grained changes in group cognition that cannot be practically assessed using traditional self-reports. Findings suggest that the performance tradeoff of group cognitive diversity is not inescapable: groups

can navigate it by modulating their levels of discursive diversity to match their changing task requirements and by having members make these individual-level adjustments in ways that are collectively congruent. Teams that develop this capacity for aligning differences can potentially achieve both high levels of creativity and seamless coordination.

References

- Aggarwal, Ishani, Anita Williams Woolley. 2019. Team creativity, cognition, and cognitive style diversity. *Management Science* **65**(4) 1586–1599.
- Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby, Michael Herron. 1996. Assessing the work environment for creativity. *Academy of Management Journal* **39**(5) 1154–1184. doi:10.5465/256995.
- Ancona, Deborah, Chee-Leong Chong. 1996. Entrainment: Pace, cycle, and rhythm in organizational behavior. *Research in Organizational Behavior* **18** 251–284.
- Ancona, Deborah, Chee-Leong Chong. 1999. Cycles and synchrony: The temporal role of context in team behavior. *Research on managing groups and teams: Groups in context, Vol. 2.* Elsevier Science/JAI Press, US, 33–48.
- Berger, Peter L., Thomas Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Anchor, New York.
- Brereton, Cannon D. M. Mabogunje A. Leifer L. J., M. F. 1996. Collaboration in design teams: How social interaction shapes the product. Henri Christiaans Nigel Cross, Kees Dorst, eds., *Analyzing design activity*. Wiley, Chichester, UK, 319–341.
- Bunderson, J. S., P. Boumgarden. 2010. Structure and learning in self-managed teams: Why "bureaucratic" teams can be better learners. *Organization Science* **21**(3) 609–624. doi:10.1287/orsc.1090.0483.
- Cannon-Bowers, Janis A, Eduardo Salas, Sharolyn Converse, Jr. Castellan, N. John. 1993. *Shared mental models in expert team decision making*. Lawrence Erlbaum Associates, Inc, 221–246.
- Carley, Kathleen M. 1997. Extracting team mental models through textual analysis. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **18**(S1) 533–558.
- Carmeli, Abraham, Meyrav Yitzack Halevi. 2009. How top management team behavioral integration and behavioral complexity enable organizational ambidexterity: The moderating role of contextual ambidexterity. *The Leadership Quarterly* **20**(2) 207–218.
- Cejka, Mary Ann, Alice H Eagly. 1999. Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin* **25**(4) 413–423.
- Chapple, Eliot D. 1974. Culture and biological man: Explorations in behavioral anthropology. *Current Anthropology* **15** 53–63. doi:10.1086/201432.
- Cheryan, Sapna, Victoria C Plaut, Caitlin Handron, Lauren Hudson. 2013. The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex Roles* **69**(1) 58–71.
- Collins, Randall. 2014. *Interaction Ritual Chains*. Princeton University Press.
- Converse, Sharolyn, JA Cannon-Bowers, E Salas. 1993. Shared mental models in expert team decision making. *Individual and Group Decision Making: Current Issues*, vol. 221. Hillsdale, 221–46.

-
- Corritore, Matthew, Amir Goldberg, Sameer B Srivastava. 2020. Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly* **65**(2) 359–394.
- Cronin, Matthew A, Laurie R Weingart, Gergana Todorova. 2011. Dynamics in groups: Are we there yet? *The Academy of Management Annals* **5**(1) 571–612.
- Cropley, Arthur. 2006. In praise of convergent thinking. *Creativity Research Journal* **18**(3) 391–404.
- Cummings, J. N. 2004. Work groups, structural diversity, and knowledge sharing in a global organization. *Management Science* **50**(3) 352–364. doi:10.1287/mnsc.1030.0134.
- De Dreu, C. K. W., M. A. West. 2001. Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology* **86**(6) 1191–1201. doi:10.1037//0021-9010.86.6.1191.
- DeChurch, Leslie A, Jessica R Mesmer-Magnus. 2010. The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology* **95**(1) 32.
- DiMaggio, Paul, Clark Bernier, Charles Heckscher, David Mimno. 2018. Interaction ritual threads: Does IRC theory apply online? *Ritual, Emotion, Violence*. Routledge, 99–142.
- Edmondson, A. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* **44**(2) 350–383.
- Eliasoph, Nina, Paul Lichterman. 2003. Culture in interaction. *American Journal of Sociology* **108**(4) 735–794. doi:10.1086/367920.
- Fleishman, Edwin A., Stephen J. Zaccaro. 1992. Toward a taxonomy of team performance functions. *Teams: Their Training and Performance*. Ablex Publishing, Westport, CT, US, 31–56.
- Foucault, Michel. 2002. *Archaeology of Knowledge*. Psychology Press.
- Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences* **115**(16) E3635–E3644. doi:10.1073/pnas.1720347115.
- Gibbs, Jack P, Walter T Martin. 1962. Urbanization, technology, and the division of labor: International patterns. *American Sociological Review* 667–677.
- Gibson, Cristina, Freek Vermeulen. 2003. A healthy divide: Subgroups as a stimulus for team learning behavior. *Administrative Science Quarterly* **48**(2) 202–239.
- Goldberg, Amir, Sameer B. Srivastava, V. Govind Manian, William Monroe, Christopher Potts. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *American Sociological Review* **81**(6) 1190–1222. doi:10.1177/0003122416671873.
- Greenwald, Anthony G, Mahzarin R Banaji. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* **102**(1) 4.

-
- Griffith, T. L., M. A. Neale. 2001. *Information processing in traditional, hybrid, and virtual teams: From nascent knowledge to transactive memory*, *Research in Organizational Behavior*, vol. 23. Jai-Elsevier Sci Bv, Amsterdam, 379–421.
- Guerin, Bernard. 1986. Mere presence effects in humans: A review. *Journal of Experimental Social Psychology* **22**(1) 38–77.
- Hackman, J. Richard. 2011. *Collaborative Intelligence: Using Teams to Solve Hard Problems*. Berrett-Koehler Publishers.
- Harkins, S.G. 1987. Social loafing and social facilitation. *Journal of Experimental Social Psychology* **23** 1–18.
- Hauser, Oliver P, Francesca Gino, Michael I Norton. 2018. Budging beliefs, nudging behaviour. *Mind & Society* **17**(1-2) 15–26.
- Healey, Mark P., Timo Vuori, Gerard P. Hodgkinson. 2015. When Teams Agree While Disagreeing: Reflexion and Reflection in Shared Cognition. *Academy of Management Review* **40**(3) 399–422. doi:10.5465/amr.2013.0154.
- Jansen, Justin JP, Gerard George, Frans AJ Van den Bosch, Henk W Volberda. 2008. Senior team attributes and organizational ambidexterity: The moderating role of transformational leadership. *Journal of Management Studies* **45**(5) 982–1007.
- Kelly, Janice, Gail Clark Futoran, Joseph E McGrath. 1990. Capacity and capability: Seven studies of entrainment of task performance rates. *Small Group Research* **21**(3) 283–314.
- Kelly, Janice R, Steven J Karau. 1993. Entrainment of creativity in small groups. *Small Group Research* **24**(2) 179–198.
- Kelly, Janice R, Joseph E McGrath. 1985. Effects of time limits and task types on task performance and interaction of four-person groups. *Journal of Personality and Social Psychology* **49**(2) 395.
- Knight, Don, Craig L Pearce, Ken G Smith, Judy D Olian, Henry P Sims, Ken A Smith, Patrick Flood. 1999. Top management team diversity, group process, and strategic consensus. *Strategic Management Journal* **20**(5) 445–465.
- Knorr Cetina, Karin, Urs Bruegger. 2002. Global microstructures: The virtual societies of financial markets. *American Journal of Sociology* **107**(4) 905–950. doi:10.1086/341045.
- Kozlowski, Austin C., Matt Taddy, James A. Evans. 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review* **84**(5) 905–949. doi: 10.1177/0003122419877135.
- Larson, Jr., James R. 2010. *In Search of Synergy in Small Group Performance*. Psychology Press, New York.
- Lewis, David. 2002. *Convention: A Philosophical Study*. Wiley.
- Liang, D., R. Moreland, L. Argote. 1995. Group versus individual training and group performance: The mediating role of transactive memory. *Personality Social Psychology Bulletin* **21** 384–393.

-
- March, James G. 1991. Exploration and exploitation in organizational learning. *Organization Science* **2**(1) 71–87. doi:10.1287/orsc.2.1.71.
- Marks, Michelle A., John E. Mathieu, Stephen J. Zaccaro. 2001. A temporally based framework and taxonomy of team processes. *Academy of Management Review* **26**(3) 356–376. doi:10.5465/amr.2001.4845785.
- Mathieu, John E., Tonia S. Heffner, Gerald F. Goodwin, Eduardo Salas, Janis A. Cannon-Bowers. 2000. The influence of shared mental models on team process and performance. *Journal of Applied Psychology* **85**(2) 273–283.
- McGrath, Joseph E. 1991. Time, interaction, and performance (TIP): A theory of groups. *Small Group Research* **22**(2) 147–174. doi:10.1177/1046496491222001.
- McGrath, Joseph E, Janice R Kelly. 1986. *Time and Human Interaction: Toward a Social Psychology of Time*. Guilford Press.
- McGrath, Joseph E, Janice R Kelly, Diane E Machatka. 1984. The social psychology of time: Entrainment of behavior in social and organizational settings. *Applied Social Psychology Annual* **5** 21–44.
- Mihalache, Oli R, Justin JP Jansen, Frans AJ Van den Bosch, Henk W Volberda. 2014. Top management team shared leadership and organizational ambidexterity: A moderated mediation framework. *Strategic Entrepreneurship Journal* **8**(2) 128–148.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., 3111–3119.
- Miller, C Chet, Linda M Burke, William H Glick. 1998. Cognitive diversity among upper-echelon executives: Implications for strategic decision processes. *Strategic Management Journal* **19**(1) 39–58.
- Milliken, Frances J, Luis L Martins. 1996. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review* **21**(2) 402–433.
- Mohammed, Susan, Lori Ferzandi, Katherine Hamilton. 2010. Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management* **36**(4) 876–910. doi:10.1177/0149206309356804.
- Moreland, Richard, Linda Argote, Ranjani Krishnan. 1998. Training people to work in groups. R. Tindale, L. Heath, eds., *Theory and Research on Small Groups*. Plenum Press, New York, 37–60.
- Moreland, Richard L., Linda Argote, Ranjani Krishnan. 1996. Socially shared cognition at work: Transactive memory and group performance. *What's Social about Social Cognition? Research on Socially Shared Cognition in Small Groups*. Sage Publications, Inc, Thousand Oaks, CA, US, 57–84. doi:10.4135/9781483327648.n3.
- O'Reilly, Charles A, Michael L Tushman. 2013. Organizational ambidexterity: Past, present, and future. *Academy of Management Perspectives* **27**(4) 324–338.

-
- Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies - New Edition*. Princeton University Press.
- Pelled, Lisa Hope, Kathleen M. Eisenhardt, Katherine R. Xin. 1999. Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative Science Quarterly* **44**(1) 1–28. doi:10.2307/2667029.
- Prince, Carolyn, Eduardo Salas. 1993. Training and research for teamwork in the military aircrew. *Cockpit Resource Management*. Academic Press, San Diego, CA, US, 337–366.
- Reagans, Ray, Ella Miron-Spektor, Linda Argote. 2016. Knowledge utilization, coordination, and team performance. *Organization Science* **27**(5) 1108–1124.
- Reagans, Ray, Ezra Zuckerman, Bill McEvily. 2004. How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly* **49**(1) 101–133. doi:10.2307/4131457.
- Reagans, Ray, Ezra W Zuckerman. 2001. Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science* **12**(4) 502–517.
- Reid, Fraser JM, Susan Reed. 2000. Cognitive entrainment in engineering design teams. *Small Group Research* **31**(3) 354–382.
- Ren, Yuqing, Linda Argote. 2011. Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *Academy of Management Annals* **5**(1) 189–229.
- Rogan, Michelle, Marie Louise Mors. 2014. A network perspective on individual-level ambidexterity in organizations. *Organization Science* **25**(6) 1860–1877.
- Salancik, G. R., J. Pfeffer. 1978. Social information-processing approach to job attitudes and task design. *Administrative Science Quarterly* **23**(2) 224–253.
- Schroeder, Juliana, Michael Kardas, Nicholas Epley. 2017. The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological Science* **28**(12) 1745–1762.
- Schwaber, Ken. 1997. SCRUM Development Process. Jeff Sutherland, Cory Casanave, Joaquin Miller, Philip Patel, Glenn Hollowell, eds., *Business Object Design and Implementation*. Springer, London, 117–134. doi:10.1007/978-1-4471-0947-1_11.
- Sørensen, Jesper B. 2002. The strength of corporate culture and the reliability of firm performance. *Administrative Science Quarterly* **47**(1) 70–91. doi:10.2307/3094891.
- Srikanth, Kannan, Sarah Harvey, Randall Peterson. 2016. A dynamic perspective on diverse teams: Moving from the dual-process model to a dynamic coordination-based model of diverse team performance. *Academy of Management Annals* **10**(1) 453–493.
- Srivastava, Sameer B, Mahzarin R Banaji. 2011. Culture, cognition, and collaborative networks in organizations. *American Sociological Review* **76**(2) 207–233.

-
- Srivastava, Sameer B., Amir Goldberg, V. Govind Manian, Christopher Potts. 2018. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science* **64**(3) 1348–1364. doi:10.1287/mnsc.2016.2671.
- ten Brinke, Leanne, Christopher C Liu, Dacher Keltner, Sameer B Srivastava. 2016. Virtues, vices, and political influence in the US Senate. *Psychological Science* **27**(1) 85–93.
- Thompson, Leigh, Gary Alan Fine. 1999. Socially shared cognition, affect, and behavior: A review and integration. *Personality and Social Psychology Review* **3**(4) 278–302. doi:10.1207/s15327957pspr0304_1.
- Tushman, Michael L, Charles A O'Reilly. 1996. Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review* **38**(4) 8–29.
- Warner, Rebecca M. 1979. Periodic rhythms in conversational speech. *Language and Speech* **22**(4) 381–396.
- Warner, Rebecca M. 1992. Cyclicity of vocal activity increases during conversation: Support for a nonlinear systems model of dyadic social interaction. *Behavioral Science* **37**(2) 128–138.
- Wegner, Daniel M. 1987. Transactive memory: A contemporary analysis of the group mind. *Theories of Group Behavior*. Springer, 185–208.
- Wegner, Daniel M, Ralph Erber, Paula Raymond. 1991. Transactive memory in close relationships. *Journal of Personality and Social Psychology* **61**(6) 923.
- Weick, Karl E. 1979. *The Social Psychology of Organizing*. Reading, MA.
- Woolley, Anita Williams, Christopher F. Chabris, Alex Pentland, Nada Hashmi, Thomas W. Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science* **330**(6004) 686–688. doi:10.1126/science.1193147.
- Wuchty, S., B. F. Jones, B. Uzzi. 2007. The increasing dominance of teams in production of knowledge. *Science* **316**(5827) 1036–1039. doi:10.1126/science.1136099.

FIGURES

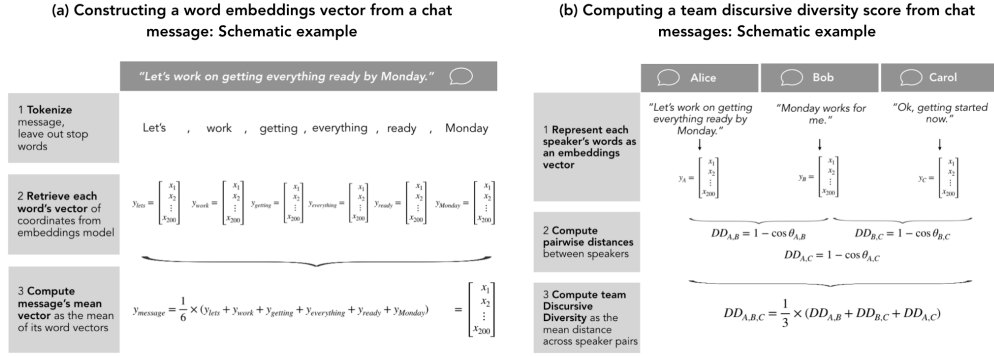


Figure 1 Procedure for Calculating Team Discursive Diversity Scores

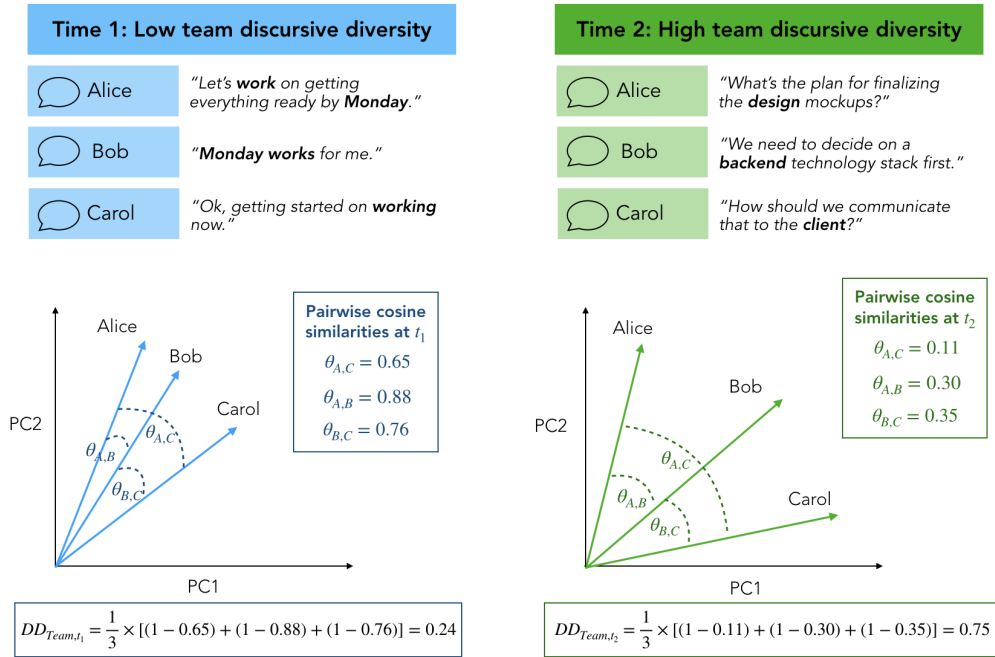
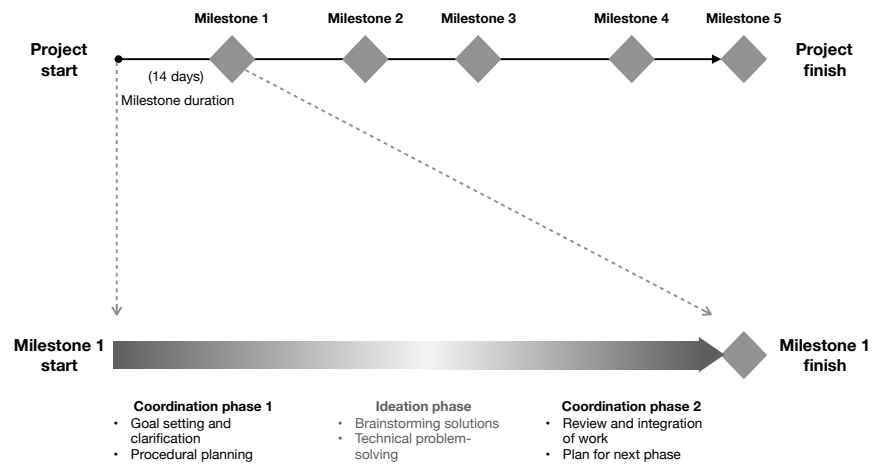
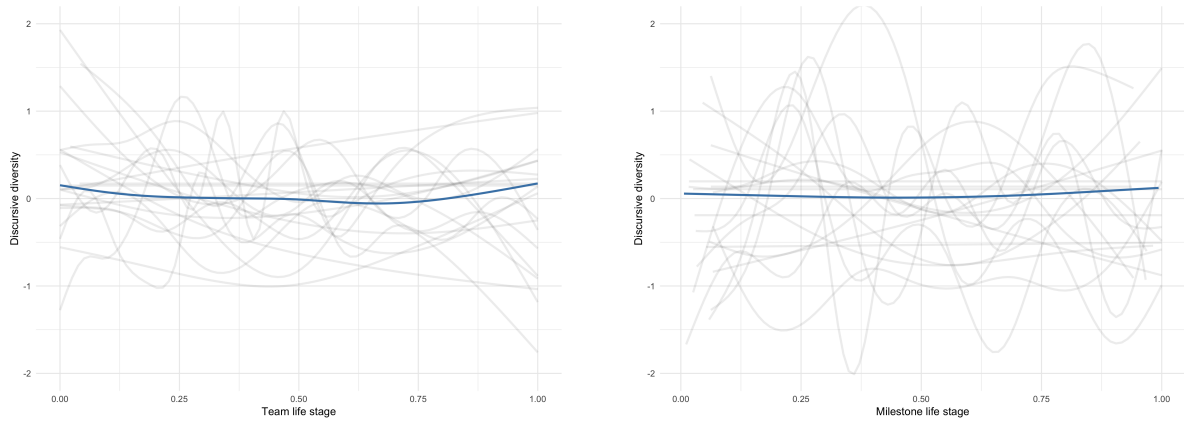


Figure 2 Schematic Illustration of Low versus High Team Discursive Diversity Scores



Note: The lower part of the figure depicts the three distinct phases that occur within each milestone, in this case Milestone 1.

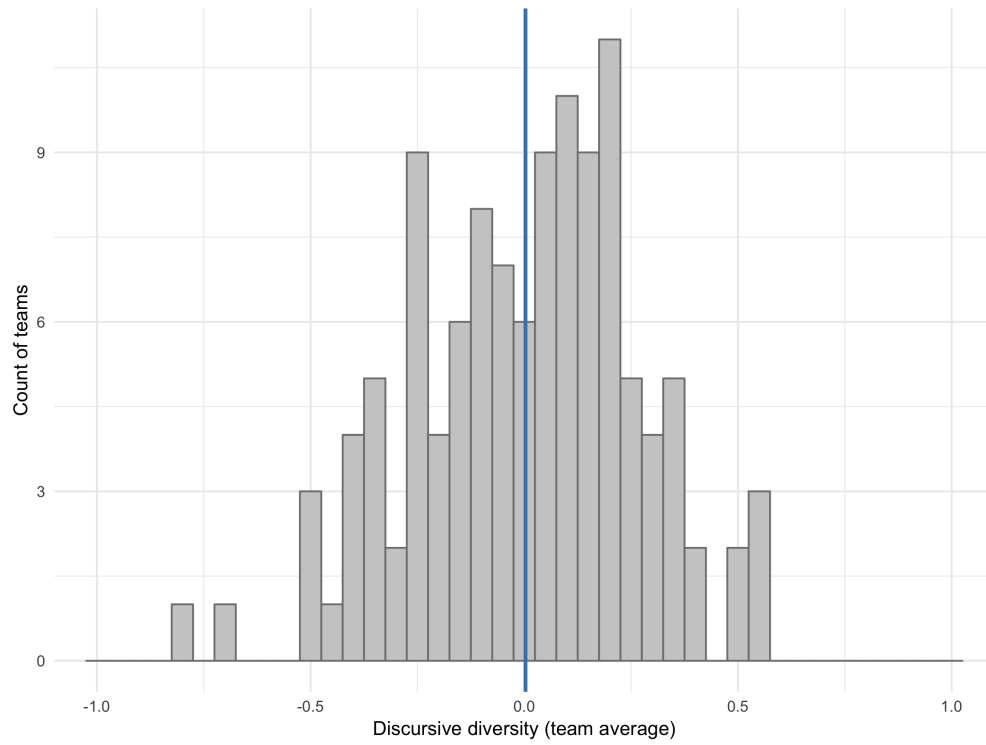
Figure 3 Schematic Representation of a Hypothetical Project that Includes Five Milestones of Varying Length



(a) Discursive diversity by team life stage for 20 randomly selected teams (b) Discursive diversity by milestone life stage for 20 randomly selected milestone phases

Note: The blue lines, respectively, represent the mean level of discursive diversity across all 117 teams at a given team life stage (Panel a) and the mean level of discursive diversity across all milestone phases for a given milestone life stage (Panel b).

Figure 4 Discursive Diversity Across Team and Milestone Life Stages



Note: The blue line represents the mean level of discursive diversity across all 117 teams.

Figure 5 Distribution of Discursive Diversity across 117 teams

TABLES

Table 1 Analogy Tasks Based on Word Embedding Model Trained on Entire Set of Teams’ Slack Archives

Analogy task	Answer
Bug - code = ?	Issue
Milestone + deliverable = ?	Sprint
Sprint - pressure = ?	Phase
Man + casual = ?	Dude
Instagram - photos = ?	Facebook
Machine - software = ?	Device
Machine + intelligent = ?	Brain
California - startup = ?	Australia
Human + desires + art = ?	Culture
Visual - creative = ?	Polish
Team - community = ?	@-tag
Man + programmer = ?	Beard
Woman + programmer = ?	Roadblock

Note: Prior research has established that gender inequality is a persistent issue in software development. This inequality stems in part from gender-stereotypic images of occupations (Cejka and Eagly 1999), including those presented in the media (Cheryan et al. 2013), that associate masculine, rather than feminine, characteristics with success in the profession. The word analogies shown in Table 1 suggest that such gender bias may also be present in the present empirical setting. This underscores a broader issue at the intersection of machine learning and the social sciences: while certain machine learning algorithms are apt at learning the semantic and cultural relationships present in the training data, they are also prone to encoding the societal biases inherent in that data (Garg et al. 2018).

Table 2 Summary Statistics and Bivariate Correlations

Summary Statistics				Bivariate correlations					
	N	Mean	St. Dev.	Success	Discursive diversity (milestone mean)	Discursive diversity (stage 1)	Discursive diversity (stage 2)	Discursive diversity (stage 3)	Incongruent modulation
Success	630	0.673	0.469						
Discursive diversity (milestone mean)	630	0.645	0.123	0.02					
Discursive diversity (stage 1)	568	0.65	0.145	-0.02	0.65***				
Discursive diversity (stage 2)	591	0.65	0.155	0.08	0.74***	0.19***			
Discursive diversity (stage 3)	572	0.636	0.137	-0.02	0.69***	0.15***	0.33***		
Incongruent modulation	630	0.09	0.093	-0.12**	0.20***	0.17***	0.15***	0.10*	
Milestone number	630	3.8	2.448	0.11*	0.04	0	0.04	0.05	-0.01

Note: *** p<0.001, ** p<0.01, * p<0.05

Table 3 Model Results for Hypotheses 1 and 2: Conditional Logistic Regressions Predicting Milestone Success

	<i>Dependent variable:</i>						
	Success						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Discursive diversity (milestone mean)	0.010 (0.079)						
Discursive diversity (stage 1)		-0.094* (0.066)			-0.121** (0.072)		-0.106* (0.073)
Discursive diversity (stage 2)			0.072+ (0.065)		0.130** (0.083)		0.155** (0.084)
Discursive diversity (stage 3)				-0.029 (0.073)	-0.116* (0.087)		-0.124* (0.087)
Incongruent modulation						-0.156** (0.076)	-0.136* (0.082)
Milestone number	0.053** (0.028)	0.065** (0.032)	0.057** (0.030)	0.069*** (0.031)	0.087*** (0.035)	0.050** (0.028)	0.085*** (0.035)
Fixed effects: Team	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects: Milestone length	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	630	568	591	572	521	630	521
R ²	0.132	0.149	0.145	0.143	0.164	0.138	0.168
<i>Note:</i>			+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001				

Models were estimated using cluster-robust standard errors. The differences in sample sizes arise from the fact that stage-wise discursive diversity is not defined for stages where only one team member sent messages, since discursive diversity requires at least two team members to send messages during a given stage.

Table 4 Model Results for Supplemental Analyses

	<i>Dependent variable:</i>				
	Success				Team discursive diversity <i>OLS</i> (5)
	(1)	<i>Conditional logistic</i>		(4)	
	(2)	(3)			
Leader-to-follower DD (stage 1)	-0.164* (0.099)		-0.203** (0.114)		
Leader-to-follower DD (stage 2)	0.169** (0.103)		0.098 (0.115)		
Leader-to-follower DD (stage 3)	-0.215** (0.116)		-0.236** (0.134)		
Between-follower DD (stage 1)		-0.070 (0.103)	0.010 (0.112)		
Between-follower DD (stage 2)		0.019 (0.118)	-0.078 (0.127)		
Between-follower DD (stage 3)		0.041 (0.115)	0.113 (0.120)		
Within-person discursive range				-0.092+ (0.053)	0.249** (0.005)
Milestone number	0.089*** (0.035)	0.065** (0.036)	0.068** (0.037)	0.051** (0.028)	-0.014 (0.606)
Fixed effects: Team	Yes	Yes	Yes	Yes	Yes
Fixed effects: Milestone length	Yes	Yes	Yes	Yes	Yes
Observations	517	482	477	631	631
R ²	0.169	0.145	0.159	0.134	0.458

Note:

+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

Models were estimated using cluster-robust standard errors.