TRACE: A Corpus for Computational Analysis of Team Creativity

Anonymous ACL submission

Abstract

Understanding how discussion dynamics shape team creativity has been limited by the difficulty of measuring process at scale. We introduce TRACE, a corpus of 309 group discussions from 103 teams (460 participants) across six creative problem-solving tasks. The dataset follows an input-process-output framework, integrating team composition (demographics, personalities), full discussion transcripts, and creativity outcomes. Using sentence embeddings and factor analysis, we identify four interpretable discussion dimensions: **Coherence**, **Exploration**, **Convergence**, and **Participation**. Analysis reveals a depth-breadth trade-off: coherent idea development inversely relates to semantic exploration. Larger teams explore more broadly but converge less effectively while team diversity shapes participation patterns more than discussion content. Novelty and usefulness in the creativity outcomes follow distinct pathways: Exploration and Convergence predict novelty, whereas Coherence predicts usefulness. These findings ground our understanding of how teams talk their way to creative solutions and provide guidance for designing multiagent systems.[1]

1 Introduction

How do teams talk their way to creative solutions? The answer matters both for understanding human collaboration and for designing multiagent systems that generate creative outputs (Tran et al., 2025; Lin et al., 2025; Barbosa et al., 2024). Yet despite decades of research on team creativity, the discussion process itself remains poorly understood. We know that team composition shapes creativity outcomes (Bell et al., 2011; Barry and Stewart, 1997) and that divergent and convergent thinking play distinct roles (Guilford, 1967; Girotra et al., 2010). What we lack is the ability to observe these processes from the perspective of natural language processing (NLP), at scale, as they unfold in real conversations.

This gap persists because measuring discussion dynamics is methodologically difficult. Standard approaches such as post-hoc surveys capture participants' perceptions rather than actual behavior (Hoever et al., 2012). Manual coding provides rich insight but limits studies to small samples—often fewer than ten groups (Harvey and Kou, 2013; Stempfle and Badke-Schaub, 2002). Experimental designs frequently impose artificial structures, separating divergent and convergent phases into distinct sessions (Coursey et al., 2019), rather than observing how teams naturally transition between exploration and focus. Recent advances in NLP offer new possibilities: NLP methods can now quantify semantic diversity, coherence, and temporal dynamics at scale. However, existing resources lack the integration needed to apply these methods to team creativity. Meeting corpora such as AMI (Carletta et al., 2005) provide dialogues but no outcome measures whereas creativity datasets include ratings but rarely release raw transcripts. No existing resource combines the three components required to study the complete Input-Process-Output (IPO) pathway computationally: individual-level team composition, discussion transcripts, and creativity outcomes.

We introduce TRACE (**T**eam **R**easoning **A**nd **C**reativity **E**xploration), a corpus designed to bridge the divide between natural discussion dynamics and objective creativity outcomes. Unlike existing resources that offer dialogues without outcomes or creativity ratings without transcripts, TRACE captures the complete IPO pathway. It integrates high-quality transcripts with detailed individual differences (personality, demographics) and outcome ratings for both novelty and usefulness.

[1] Our corpus and code will be released upon acceptance.

This design enables computational operationalization of discussion dynamics. Applying factor analysis to embedding-based discourse features, we demonstrate how NLP methods can recover theoretically meaningful discussion dimensions: COHERENCE, EXPLORATION, CONVERGENCE, and PARTICIPATION, and quantify how they interact with team composition to drive creative performance.

Our contributions include:

1. TRACE, a dataset of 309 group discussions from 103 teams (460 participants) solving six creative tasks, totaling 28.6 hours of interaction with reliable creativity scores.
2. A computational framework for operationalizing discussion process using NLP methods, uncovering four interpretable dimensions that capture theoretically meaningful constructs.
3. A thorough empirical characterization of IPO pathways. We demonstrate that the discussion dynamics required for novelty (broad exploration) are statistically distinct from those required for usefulness (coherent convergence). Furthermore, we show how structural factors like team size force trade-offs between these processes, providing empirical grounding for theories of collective creativity.

By quantitatively evaluating discussion process through computational linguistic analysis and linking it to team composition and creativity outcomes, TRACE provides a data-driven foundation for understanding how creative ideas emerge through collaborative interaction.

2 Related Work

Team Creativity Research A substantial body of research has examined how team composition relates to creativity outcomes. Meta-analytic evidence reveals inconsistent effects: demographic diversity varies by dimension and criterion (Bell et al., 2011), with deep-level differences more consistently linked to novelty than surface-level variation (Wang et al., 2019). Aggregate characteristics such as mean openness and team size predict outcomes with modest effect sizes (Barry and Stewart, 1997; Wu et al., 2019).

Prior research has also investigated process mechanisms that unfold during creative collaboration. For example, empirical work has examined how teams transition between phases of divergent thinking (generating multiple options) and convergent thinking (selecting and refining the best) (Girotra et al., 2010). Three process constructs have emerged as particularly important: *information elaboration*, the extent to which teams build on each other's contributions, predicts outcomes more strongly than idea quantity (Van Knippenberg et al., 2004; Coursey et al., 2019); *evaluation* during ideation determines which ideas survive (Harvey and Kou, 2013); and *participation balance* influences whether diverse perspectives are actually expressed (Woolley et al., 2010).

Operationalizing these constructs has proved challenging. Many studies experimentally impose discussion structures rather than observing natural dynamics (Coursey et al., 2019). Others rely on post-hoc surveys assessing subjective perceptions rather than observed behavior (Hoever et al., 2012), introducing retrospective bias. Qualitative coding of actual discourse provides insight but is only feasible with limited sample sizes (Stempfle and Badke-Schaub, 2002). These constraints have restricted understanding of team creativity as a temporal, interactional phenomenon.

Computational Approaches to Creativity Computational linguistics offers tools to address these limitations. Early work established that features of creative products, such as lexical compositions and semantic associations, could be quantified using information-theoretic and distributional measures (Kuznetsova et al., 2013). More recent studies apply NLP methods to evaluate creative outputs at scale, using semantic distance to score novelty (Beaty and Johnson, 2021) or training models to align with human ratings (Luchini et al., 2025; Li et al., 2025). However, this line of research conceptualizes creativity primarily as a property of final artifacts rather than of the processes that generate them.

A related strand operationalizes creativity-relevant cognitive processes in controlled settings. Verbal fluency tasks, for instance, model switching and clustering dynamics using language models (Zarrieß et al., 2025). While such predefined features capture aspects of individual ideation, they do not encompass the open-ended dynamics of real team discussions, nor connect these to collaborative outcomes.

Fewer studies apply computational methods

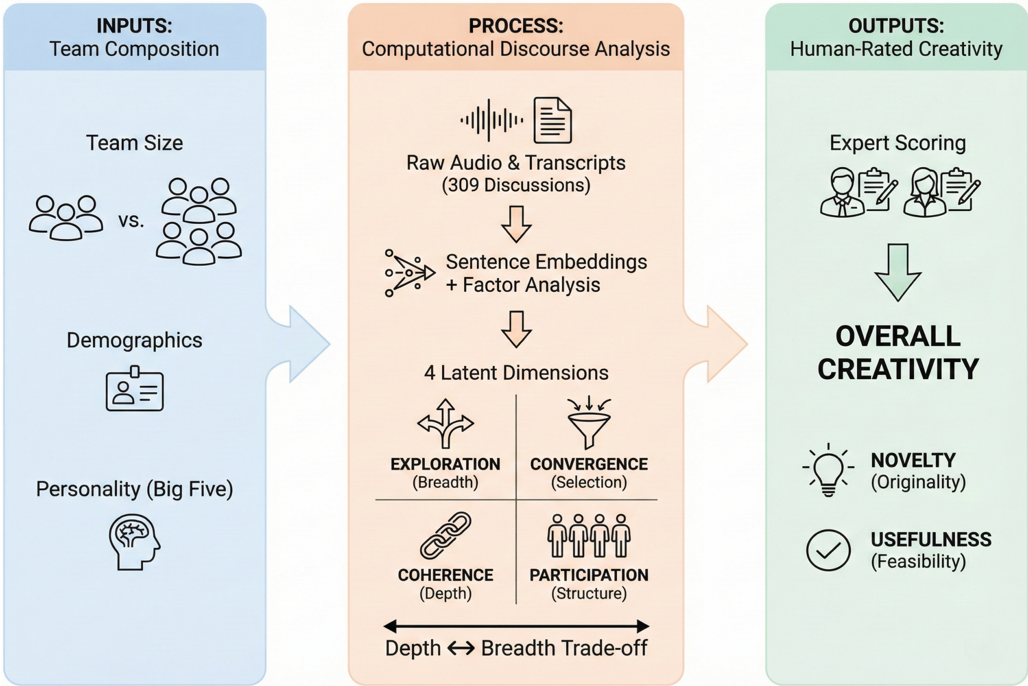


Figure 1: Overview of the TRACE corpus and analysis framework. Team composition variables (inputs) are linked to discussion transcripts analyzed through computational linguistics methods, revealing four latent discussion dimensions. These process patterns predict creativity outcomes (outputs) rated separately on novelty and usefulness.

directly to creative discourse. Work on group communication analysis identifies dimensions including participation balance and semantic novelty (Dowell et al., 2019; Huber et al., 2019), suggesting that computational approaches can recover meaningful constructs. However, they often lack integration with creativity outcome measures, failing to test process–outcome relationships within a complete IPO framework.

Existing Corpora and the Gap There is a resource gap among the widely used meeting corpora. For example, AMI (Carletta et al., 2005) and ICSI (Janin et al., 2003) provide rich dialogue annotations but omit creativity outcomes and individual participant characteristics. More recent corpora (Hu et al., 2023; Schroeder et al., 2024; Cai et al., 2025) similarly focus on transcripts without any link to task performance. Creativity-focused datasets typically include outcome ratings but rarely release full transcripts. No existing resource integrates the three components needed to study IPO computationally: individual-level team composition, complete discourse transcriptions, and creativity outcomes with separate novelty and usefulness ratings. TRACE addresses this gap by providing an integrated corpus designed specifically to enable computational analysis of how team characteristics shape discussion dynamics, and how those dynamics predict creative performance.

3 Corpus

This section describes the data collection procedure, transcription pipeline, and creativity outcome evaluation.

3.1 Data Collection

Participants were recruited through email invitations and an online research participant management system. Eligibility criteria required participants to be at least 18 years old and fluent in English. Participants received £15 for approximately one hour of participation, which included survey completion and three group discussion tasks.

We recruited 342 groups, each composed of three or six participants, to engage in collaborative creative problem-solving tasks. After excluding sessions with incomplete survey responses or recording failures, 309 group discussions (103 unique groups, 460 participants) were retained for analysis.

Each group completed one of two task sets comprising three problems, spanning general social challenges (e.g., strategies to reduce plastic waste), specialized organizational issues (e.g., supply chain optimization), and hypothetical scenarios (e.g., a pandemic transmitted by saying "sorry"). Groups were allotted up to 12 minutes per task and required to submit a single solution.

Following each session, participants completed a survey capturing demographic variables (e.g., age, gender, ethnicity, education, employment, work experience, professional area, and discipline), Big Five personality profiles, and additional information about perceived creativity, time and size adequacy, effort, self-identified roles, and prior familiarity with teammates.

3.2 Transcription Pipeline

Discussions were transcribed using Microsoft Azure's speech-to-text service with speaker diarization. Speaker count was specified based on the known group size. The resulting transcripts provide utterance-level segmentation with timestamps, speaker attribution, and turn duration.

To validate transcription quality, we sampled 12 discussions (1,501 utterances) stratified by group size for manual verification by two research assistants. Each utterance was annotated for semantic accuracy (whether the transcribed content correctly captured the spoken words) and speaker diarization accuracy (whether the utterance was assigned to the correct speaker). Semantic accuracy reached 93.34%, and speaker diarization accuracy reached 89.74%. These accuracy levels are sufficient for computational feature extraction, as aggregate discourse patterns are robust to individual transcription errors.

3.3 Creativity Outcomes

All solutions submitted by the groups were independently evaluated by five trained human judges on novelty (originality relative to other solutions) and usefulness (practicality and feasibility) using a 10-point scale, following the Consensual Assessment Technique (Amabile, 1982). To provide a broader reference distribution, judges rated both human-generated solutions from the current dataset and LLM-generated solutions from a separate study (a total of approximately 850 solutions for each task).

Before full annotation, judges first rated a pilot set of 80 solutions for each task and then participated in a calibration session to establish shared anchors for each scale point before rating the whole set of solutions. Inter-rater reliability among the human judges was assessed using the intraclass correlation coefficient (ICC), yielding satisfactory agreement for both novelty and usefulness (Novelty: Mean ICC = 0.83, minimum 0.73; Usefulness: Mean ICC = 0.75, minimum 0.70).

To account for task-specific difficulty, novelty and usefulness scores were min–max normalized within each task. A composite creativity score was computed as the product of normalized novelty and usefulness, consistent with theoretical accounts treating creativity as jointly constituted by these two dimensions (Cropley, 2025; Runco and Jaeger, 2012; Simonton, 2012).

3.4 Descriptive Statistics

TRACE contains 309 discussions contributed by 103 teams, comprising 36,223 utterances and covering 28.6 hours of recorded interaction. Each team completed three tasks drawn from one of two task sets. Task Set 1 (Plastic Waste, Supply Chain, Pandemic) contains 45 discussions per task (135 total), while Task Set 2 (Education Inequality, Employee Attrition, Singing) contains 58 discussions per task (174 total). Discussion lengths averaged 5.5 minutes ($SD = 2.4$), ranging from 0.5 to 10.7 minutes.

The dataset comprises a heterogeneous participant pool spanning diverse demographic, educational, and professional backgrounds. Team composition varies widely across groups, encompassing various combinations of demographic and background attributes, including both homogeneous and heterogeneous configurations (see Appendix A.2). This diversity of team composition enables systematic analysis of how different compositional patterns relate to interaction processes and creativity outcomes.

4 Analysis

We address three research questions within the Input-Process-Output framework: (1) What patterns characterize creativity outcomes, and what do they suggest about process? (2) How can computational methods operationalize discussion dynamics? (3) How do team composition and discussion process relate to creativity outcomes?

4.1 Creativity Outcomes

Submitted solutions were rated on novelty ($M = 0.41$, $SD = 0.20$) and usefulness ($M = 0.64$,

$SD = 0.19$), with composite creativity computed as their product after min-max normalization.

Novelty-Usefulness Trade-off. Figure 2a reveals a significant negative correlation between novelty and usefulness ($r = -0.42$, $p < .001$). This trade-off suggests that achieving high novelty often comes at the cost of practical feasibility, reflecting an inherent tension between divergent and convergent aspects of creativity.

Group Size Effect. As shown in Figure 2b, smaller teams tended to produce more novel ideas, with 3-person teams outperforming 6-person teams in novelty. In contrast, usefulness ratings showed a modest advantage for larger teams.

Task Variation. Creativity outcomes varied substantially across task types (Figure 2c). Open-ended and expressive tasks (e.g., Singing, Pandemic) yielded higher overall creativity, driven primarily by elevated novelty, whereas more constrained problem-solving tasks (e.g., Plastic Waste, Supply Chain) exhibited lower creativity scores.

4.2 Computational Operationalization of Discussion Process

A critical methodological challenge in team creativity research is measuring what happens *during* discussions. Prior work has relied on post-hoc surveys (Hoever et al., 2012) or labor-intensive manual coding, treating the discussion process as a "black box" (Kurtzberg and Amabile, 2001). Here we develop a computational framework using NLP techniques to quantitatively extract interpretable discourse features.

4.2.1 Preprocessing

Raw transcripts contain brief utterances (e.g., Yeah, OK) carrying minimal semantic content. We merged consecutive same-speaker utterances into *speaker segments*, reducing the corpus from 36,220 utterances to 22,220 segments. After filtering segments with duration <2 seconds or fewer than 5 words, 10,534 substantive segments remained for analysis.

4.2.2 Feature Extraction

We extracted 18 discourse features operationalizing four theoretically-motivated dimensions from creativity and team research: *Exploration* (semantic breadth and diversity), *Convergence* (focusing toward solutions over time), *Development* (elaborative building on ideas), and *Participation* (distribution of speaking time). Features were computed using sentence embeddings (all-MiniLM-L6-v2) and established metrics from creativity research (Beaty and Johnson, 2021; Guilford, 1967) and team dynamics literature (Woolley et al., 2010; Van Knippenberg et al., 2004). Full feature definitions are provided in Appendix A.3.

4.2.3 Pattern Discovery via Factor Analysis

To identify latent discussion patterns, we conducted exploratory factor analysis (EFA). Bartlett's test ($\chi^2 = 2993.5$, $p < .001$) and KMO (0.70) confirmed factorability. Parallel analysis indicated four factors explaining 53.1% of variance. We retained 12 features with absolute loadings greater than 0.5 for subsequent modeling (see Appendix Table 5).

Table 1 summarizes the identified patterns. We interpreted these factors by examining the features loading onto each of these constructs:

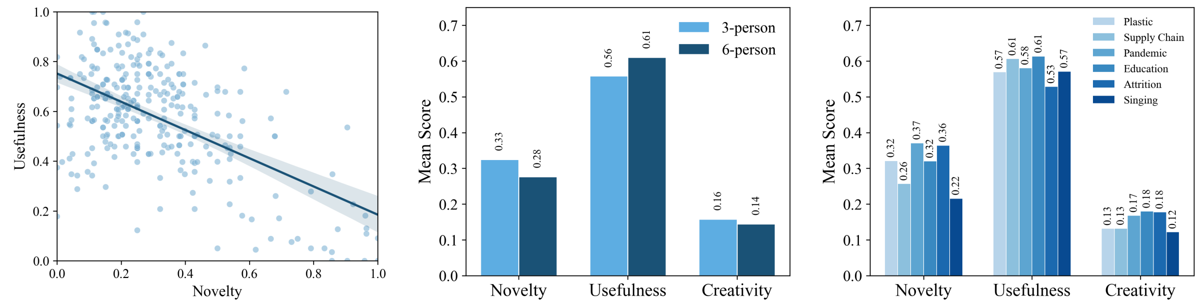
Coherence (originally Development). In addition to elaboration features that loaded positively, *semantic diversity* showed a strong *negative* loading ($-.88$). This reveals a depth-breadth trade-off: maintaining thematic focus (coherence) empirically comes at the cost of broad exploration. We labeled this factor COHERENCE to reflect this extension.

Exploration. This factor combined semantic breadth indicators with *response latency* (negative loading). This integration suggests that active pace facilitates exploration, where conceptually diverse phases are empirically linked to faster-paced interactions.

Convergence and Participation. These factors emerged as distinct constructs consistent with our initial operationalization, dominated by diversity trends and speaker entropy measures, respectively.

4.3 Team Composition and Discussion Dynamics

We adopted a two-stage analytical strategy to examine how team composition shapes discussion process. First, we performed ordinary least squares (OLS) regression with task-fixed effects to identify how structural features (team size, aggregate diversity) predict discussion patterns across different problem domains. Second,



(a) Novelty vs. Usefulness (b) Scores by team size (c) Scores by task type

Figure 2: Creativity outcomes across the TRACE dataset.

Table 1: Latent discussion patterns identified through exploratory factor analysis (EFA).

Pattern	Var.	Key features
COHERENCE	18.3	revisit score (.69); local coherence (.69); final idea alignment (.51); semantic div (−.88)
EXPLORATION	15.4	effective dim (.80); semantic div_{max} (.64); response latency (−.67); cluster dispersion (−.55)
CONVERGENCE	9.8	convergence ratio (.85); diversity trend (−.82)
PARTICIPATION	9.5	speaker entropy (.75); dominance ratio (−.98)

Note. Values in parentheses indicate standardized factor loadings. Negative signs indicate inverse relationships with the latent factors.

Table 2: OLS regression results: Team composition predicting discussion patterns.

Predictor	C	E	V	P
Team size	$-.10^{***}$	$+.11^{***}$	$-.05^{*}$	$+.11^{***}$
Background div.	$-.14$	-6.88^{***}	-1.40	$+8.36^{***}$
Personality div.	$+.15$	-3.94^{***}	-0.67	$+3.15^{***}$
Overall div.	$+.16$	$+10.72^{***}$	$+1.99$	-11.62^{***}

Note. Columns correspond to discussion patterns: C = Coherence, E = Exploration, V = Convergence, P = Participation. Standardized coefficients reported. Task-fixed effects included in all models. $^{***}p < .001$, $^{*}p < .05$.

we used correlation analysis to unpack the specific attribute-level mechanisms (e.g., specific personality traits or background attributes) driving these relationships.

4.3.1 Structural Predictors

Table 2 presents the regression results controlling for task effects. We observed two core structural relationships:

The Breadth-Focus Trade-off of Team Size. Team size proved to be a double-edged sword. Larger teams exhibited significantly higher EXPLORATION ($\beta = .11$) and PARTICIPATION ($\beta = .11$), but lower COHERENCE ($\beta = -.10$) and CONVERGENCE ($\beta = -.05$). This indicates that while adding members expands semantic territory and promotes egalitarian interaction, it imposes coordination costs that hinder thematic focus.

Diversity Drives Participation. The most robust effect of diversity was on interaction dynamics rather than semantic content. Both background diversity ($\beta = 8.36$) and personality diversity ($\beta = 3.15$) strongly predicted higher PARTICIPATION (entropy). This suggests that heterogeneous teams are structurally less prone to dominance by a single speaker, fostering more balanced turn-taking dynamics.

Correlation analysis revealed attribute-level mechanisms underlying these structural effects. For example, work experience and age are linked to EXPLORATION, while Agreeableness is related to COHERENCE (see Appendix Table 9).

4.4 Process Patterns and Creativity Outcomes

To evaluate how discussion dynamics impact team performance, we conducted a series of OLS regressions to model three distinct dependent variables: overall *Creativity* and its constituent dimensions, *Novelty* and *Usefulness*. The independent variables were the four factor scores derived from the EFA model. To ensure robustness, we controlled for task-fixed effects in the analysis to rule out problem-specific variation.

Table 3 summarizes the regression results.

4.4.1 Predictors of Overall Creativity

Two process patterns emerged as significant drivers of overall creativity. **EXPLORATION** was a positive predictor ($\beta = .016, p <$

Table 3: OLS regression results: Discussion patterns predicting creativity outcomes.

Pattern	Creativity	Novelty	Usefulness
COHERENCE	$+.004$	$-.009$	$+.027^{\dagger}$
EXPLORATION	$+.016^{*}$	$+.033^{*}$	$-.001$
CONVERGENCE	$+.014^{**}$	$+.033^{*}$	$+.006$
PARTICIPATION	$-.004$	$-.008$	$+.015$

$^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$. Task-fixed effects included.

.05), supporting divergent thinking theory where broader semantic associations yield more creative outcomes. Similarly, **CONVERGENCE** showed a significant positive effect ($\beta = .014, p < .01$). This dual significance finding suggests that high-performing teams successfully manage the tension between expanding the problem space (exploration) and focusing attention on solutions (convergence).

4.4.2 Differential Effects on Novelty and Usefulness

Decomposing creativity into sub-dimensions revealed a functional specialization of these discussion patterns.

NOVELTY was exclusively driven by the semantic breadth and focus dynamics. Both EXPLORATION and CONVERGENCE significantly predicted novelty ratings ($\beta = .033, p < .01$).

USEFULNESS, in contrast, showed a distinct profile. It was unrelated to EXPLORATION but exhibited a marginal positive association with COHERENCE ($\beta = .027, p < .10$). This reflects the **depth-breadth trade-off** identified in our structural analysis: the sustained, focused elaboration (Coherence) required to ensure a practical and useful solution appears to compete with the attempts to associate remote concepts required for novelty.

5 Discussion

5.1 Computational Measurement of Discussion Process

A persistent challenge in team creativity research is operationalizing what happens during team discussions. Prior work has either relied on experimentally imposed phase structures, post-hoc self-reports of discussion process, or labor-intensive manual coding schemes that limit scalability. Our work demonstrates that NLP methods can provide scalable, reproducible, and theoretically interpretable measures of discussion process.

Factor analysis reveals underlying patterns. COHERENCE exhibits inverse loadings for idea development and semantic exploration—a depth-breadth trade-off consistent with Lu et al. (2025)'s fNIRS evidence for neurally distinct ideation pathways (flexibility, persistence, convergence). Response latency loads onto EXPLORATION, linking semantically diverse discussions to faster-paced interaction; this parallels Harada (2020)'s reinforcement learning framework where exploration behaviors predicted divergent thinking. Rapid turn-taking may function as behavioral exploration, preventing premature convergence. These emergent patterns demonstrate the value of combining theory-driven features with data-driven discovery. Prior work has shown computational approaches can recover meaningful discussion constructs (Dowell et al., 2019); our contribution extends this by linking such dimensions to creativity outcomes within an IPO framework.

5.2 Theoretical Implications for Team Creativity

Our findings offer empirical grounding for understanding how team characteristics shape creativity outcomes through observable discussion dynamics.

Team size and diversity shape discussion through distinct mechanisms. Team size proved double-edged: larger teams exhibited higher EXPLORATION and PARTICIPATION but lower COHERENCE and CONVERGENCE. This aligns with meta-analytic evidence that larger teams benefit from greater human capital but suffer coordination losses (Bernerth et al., 2023); Osorio and Bornmann (2021) offer a complementary explanation via credit-sharing incentives. Our finding that smaller teams showed higher CONVERGENCE suggests they may be better positioned to develop breakthrough ideas to completion.

Diversity, in contrast, shaped participation dynamics more than semantic content, consistent with the meta-analysis showing no significant relationship between demographic diversity and information elaboration (Traylor et al., 2024). Vedres and Vásárhelyi (2023) similarly found that diversity without inclusion does not contribute to creativity. Correlation analysis revealed that certain attributes drive specific patterns: work experience and age were associated with EXPLORATION, while Agreeableness with COHERENCE, consistent with evidence that personality composition affects participation patterns (Hundschell et al., 2022; Bai et al., 2024).

Novelty and usefulness in the creativity outcomes follow distinct process pathways. Discussion patterns predict novelty and usefulness through different, sometimes opposing, mechanisms. EXPLORATION and CONVERGENCE predicted novelty, while COHERENCE marginally predicted usefulness. This differential pattern aligns with neuroscientific evidence that the Default Mode Network encodes originality while the Executive Control Network encodes adequacy (Moreno-Rodriguez et al., 2025; Yeo et al., 2024). At the team level, Sun et al. (2016) found that constructive controversy mediates effects on novelty but not usefulness. Orwig et al. (2025) also showed that future-oriented language characterized novelty evaluation and past-oriented language characterized usefulness evaluation.

The depth-breadth trade-off. The inverse relationship between COHERENCE and semantic diversity suggests teams face a fundamental strategic choice: exploring broadly sacrifices coherence development, while developing deeply sacrifices exploratory breadth. Baruah et al. (2021)'s experiment demonstrated this trade-off is process-dependent, with an entrainment effect where initial focus sets the trajectory for subsequent development. Malaie et al. (2024) provided a cognitive explanation by linking creativity to evolutionarily ancient foraging mechanisms. This trade-off has methodological implications: as Lloyd-Cox et al. (2022) found, novelty contributes more to creativity evaluation for divergent tasks while usefulness contributes more for implementation contexts. Studies operationalizing creativity as a unitary construct may obscure process effects that operate in opposite directions for its constituent dimensions.

5.3 Implications for Future Research

Beyond organizational behavior and creativity research, TRACE supports important NLP research directions. For *discourse coherence modeling*, the inverse relationship between local coherence and semantic diversity suggests that high coherence is not universally desirable; systems optimizing for creative or exploratory dialogues may benefit from reduced coherence constraints. For *semantic similarity evaluation*, our findings demonstrate that embedding-based diversity metrics predict meaningful task outcomes, extending these methods from product assessment to process analysis. For *multi-agent LLM systems*, the depth-breadth trade-off implies an architectural choice: agents optimized for coherent elaboration may produce more useful but less novel outputs than agents optimized for semantic exploration. TRACE provides empirical grounding for these design decisions. The accompanying feature extraction pipeline offers a validated framework for operationalizing discussion dynamics in collaborative AI research.

6 Conclusion

This work demonstrates that discussion dynamics, long treated as a methodological black box in team creativity research, can be computationally operationalized. The four dimensions we identified through factor analysis uncover constructs that prior work could only measure through labor-intensive coding or post-hoc surveys, and they do so at scale.

Our findings carry practical implications. The depth-breadth trade-off suggests that teams face a genuine strategic choice: exploring broadly or developing coherently. The distinct predictors of novelty and usefulness imply that interventions targeting one may not benefit the other. For multiagent system design, these results suggest that agents optimized for coherent elaboration will produce different outputs than agents optimized for semantic exploration.

By demonstrating that computational discourse features can operationalize discussion process underlying team creativity, this work provides methodological grounding for both organizational research and the design of multiagent creative systems.

Limitations

While TRACE provides a rich resource for process-level analysis of team creativity, several opportunities remain for further data enrichment. First, our computational approach focuses on semantic properties of discussion transcripts, such as embedding-based diversity, coherence, and convergence measures. However, the same semantic content, depending on conversational context, may serve different functions. Team creativity theories often emphasize the *functional roles* of utterances, for instance, whether contributions propose new ideas, elaborate on previous suggestions, evaluate feasibility, or coordinate group process (Huber et al., 2019; Bales, 1950). Future work could expand TRACE by incorporating dialogue act or discourse function annotations, enabling more direct operationalization of constructs such as idea generation, idea evaluation, and idea elaboration. Such annotations would also support investigation of sequential patterns, for instance, whether proposal-elaboration-evaluation sequences predict different outcomes than proposal-evaluation-elaboration sequences.

Second, our analyses rely exclusively on transcribed speech. Nonverbal signals, including prosodic cues (pitch, intensity, speech rate), temporal dynamics (pauses, overlaps, response latency at finer granularity), and visual signals (gaze, gesture, posture) contain important information related to the coordination, engagement, and influence in creative teams (Tsai et al., 2012). The original audio recordings (provided that proper anonymization and participant consent are in place) could support the extraction of paralinguistic features in future work.

Ethical Considerations

This study was reviewed and approved by the authors' institutional review board.

All participants provided informed consent prior to participation. The consent procedure explicitly described: (1) the purpose of the research (studying creative problem-solving in teams), (2) that audio recordings would be captured and transcribed, (3) how the data would be used and shared (for academic research, with anonymized data potentially released publicly), (4) that participation was voluntary and could be withdrawn at any time without penalty, and (5) data protection and anonymization procedures.

To protect participant privacy, all discussion transcripts and survey data were anonymized prior to analysis. Personally identifiable information was removed, and analyses were conducted at the individual or team level without attempting to infer or disclose participant identities. Any shared resources derived from this work contain only de-identified data or aggregated representations and do not include raw identifiable transcripts.

The study includes self-reported demographic attributes, personality traits, and creativity-related assessments. These measures are used solely for research purposes and should not be interpreted as diagnostic, evaluative, or predictive of individual ability, performance, or suitability in real-world settings.

Finally, while computational analysis of team discussions can advance the scientific understanding of collective creativity, we caution against the use of the proposed methods for surveillance, automated evaluation, or high-stakes decision-making about individuals or teams without appropriate safeguards, transparency, and human oversight.

Generative AI tools were used for language rephrasing and visualization assistance.

References

Teresa M Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology*, 43(5):997.

Yiming Bai, Ying Hu, Zihan Zhou, Xing Du, Yunxiang Shi, and Lisi You. 2024. Rise above prejudice against personality: Association with personality and interactive collaboration in team creativity performance. *Thinking Skills and Creativity*, 52:101539.

Robert F. Bales. 1950. *Interaction Process Analysis: A Method for the Study of Small Groups*. Addison-Wesley, Cambridge, MA.

Ricardo Barbosa, Ricardo Santos, and Paulo Novais. 2024. Collaborative problem-solving with llm: a multi-agent system approach to solve complex tasks using autogen. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 203–214. Springer.

Bruce Barry and Greg L Stewart. 1997. Composition, process, and performance in self-managed groups: the role of personality. *Journal of Applied psychology*, 82(1):62.

Jonali Baruah, Paul B Paulus, and Nicholas W Kohn. 2021. The effect of the sequence of creative processes on the quality of the ideas: The benefit of a simultaneous focus on originality and feasibility. *The Journal of Creative Behavior*, 55(4):946–961.

Roger E Beaty and Dan R Johnson. 2021. Automating creativity assessment with semdis: An open platform for computing semantic distance. *Behavior research methods*, 53(2):757–780.

Suzanne T Bell, Anton J Villado, Marc A Lukasik, Larisa Belau, and Andrea L Briggs. 2011. Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of management*, 37(3):709–743.

Jeremy B Bernerth, Jeremy M Beus, Catherine A Helmuth, and Terrance L Boyd. 2023. The more the merrier or too many cooks spoil the pot? a meta-analytic examination of team size and team effectiveness. *Journal of Organizational Behavior*, 44(8):1230–1262.

Jon Cai, Brendan King, Peyton Cameron, Susan Windisch Brown, Miriam Eckert, Dananjay Srinivas, George Baker, V Kate Everson, Martha Palmer, James H Martin, and 1 others. 2025. In search of the lost arch in dialogue: A dependency dialogue acts corpus for multi-party dialogues. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20135–20149.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, and 1 others. 2005. The ami meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction*, pages 28–39. Springer.

Lauren E Coursey, Ryan T Gertner, Belinda C Williams, Jared B Kenworthy, Paul B Paulus, and Simona Doboli. 2019. Linking the divergent and convergent processes of collaborative creativity: The impact of expertise levels and elaboration processes. *Frontiers in Psychology*, 10:699.

David H Cropley. 2025. "the cat sat on the…?" why generative ai has limited creativity. *The Journal of Creative Behavior*, 59(4):e70077.

Nia MM Dowell, Tristan M Nixon, and Arthur C Graesser. 2019. Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior research methods*, 51(3):1007–1041.

Karan Girotra, Christian Terwiesch, and Karl T Ulrich. 2010. Idea generation and the quality of the best idea. *Management science*, 56(4):591–605.

Joy P Guilford. 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1):3–14.

Tsutomu Harada. 2020. The effects of risk-taking, exploitation, and exploration on creativity. *PloS one*, 15(7):e0235698.

Sarah Harvey and Chia-Yu Kou. 2013. Collective engagement in creative tasks: The role of evaluation in the creative process in groups. *Administrative science quarterly*, 58(3):346–386.

Inga J Hoever, Daan Van Knippenberg, Wendy P Van Ginkel, and Harry G Barkema. 2012. Fostering team creativity: perspective taking as key to unlocking diversity's potential. *Journal of applied psychology*, 97(5):982.

Yebowen Hu, Timothy Ganter, Hanieh Deilamsalehy, Franck Dernoncourt, Hassan Foroosh, and Fei Liu. 2023. Meetingbank: A benchmark dataset for meeting summarization. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16409–16423.

Bernd Huber, Stuart Shieber, and Krzysztof Z Gajos. 2019. Automatically analyzing brainstorming language behavior with meeter. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–17.

Andreas Hundschell, Stefan Razinskas, Julia Backmann, and Martin Hoegl. 2022. The effects of diversity on creativity: A literature review and synthesis. *Applied Psychology*, 71(4):1598–1634.

Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and 1 others. 2003. The icsi meeting corpus. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, volume 1, pages I–I. IEEE.

Dan R Johnson, James C Kaufman, Brendan S Baker, John D Patterson, Baptiste Barbot, Adam E Green, Janet van Hell, Evan Kennedy, Grace F Sullivan, Christa L Taylor, and 1 others. 2023. Divergent semantic integration (dsi): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*, 55(7):3726–3759.

Terri R Kurtzberg and Teresa M Amabile. 2001. From guilford to creative synergy: Opening the black box of team-level creativity. *Creativity Research Journal*, 13(3-4):285–294.

Polina Kuznetsova, Jianfu Chen, and Yejin Choi. 2013. Understanding and quantifying creativity in lexical composition. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1246–1258.

Ruizhe Li, Chiwei Zhu, Benfeng Xu, Xiaorui Wang, and Zhendong Mao. 2025. Automated creativity evaluation for large language models: A reference-based approach. *arXiv preprint arXiv:2504.15784*.

Yi-Cheng Lin, Kang-Chieh Chen, Zhe-Yan Li, Tzu-Heng Wu, Tzu-Hsuan Wu, Kuan-Yu Chen, Hung-yi Lee, and Yun-Nung Chen. 2025. Creativity in llm-based multi-agent systems: A survey. *arXiv preprint arXiv:2505.21116*.

James Lloyd-Cox, Alan Pickering, and Joydeep Bhattacharya. 2022. Evaluating creativity: How idea context and rater personality affect considerations of novelty and usefulness. *Creativity Research Journal*, 34(4):373–390.

Kelong Lu, Xinyue Wang, Xinuo Qiao, Zhenni Gao, and Ning Hao. 2025. Group creativity emerges from triple ideation pathways: neurobehavioral evidence from an fnirs hyperscanning study. *Cerebral Cortex*, 35(5):bhaf129.

Simone A Luchini, Nadine T Maliakkal, Paul V DiStefano, Antonio Laverghetta Jr, John D Patterson, Roger E Beaty, and Roni Reiter-Palmon. 2025. Automated scoring of creative problem solving with large language models: A comparison of originality and quality ratings. *Psychology of Aesthetics, Creativity, and the Arts*.

Soran Malaie, Michael J Spivey, and Tyler Marghetis. 2024. Divergent and convergent creativity are different kinds of foraging. *Psychological Science*, 35(7):749–759.

Sarah Moreno-Rodriguez, Benoît Béranger, Emmanuelle Volle, and Alizée Lopez-Persem. 2025. The human reward system encodes the subjective value of ideas during creative thinking. *Communications Biology*, 8(1):37.

Bernard A Nijstad and Wolfgang Stroebe. 2006. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review*, 10(3):186–213.

Lucie Nikoleizig, Stefan C Schmukle, Maurin Griebenow, and Sascha Krause. 2021. Investigating contributors to performance evaluations in small groups: Task competence, speaking time, physical expressiveness, and likability. *PloS One*, 16(6):e0252980.

William Orwig, Roger E Beaty, Mathias Benedek, and Daniel L Schacter. 2025. Creative evaluation: The role of memory in novelty & effectiveness judgements. *Creativity research journal*, 37(3):514–522.

Antonio Osorio and Lutz Bornmann. 2021. On the disruptive power of small-teams research. *Scientometrics*, 126(1):117–133.

Mark A Runco and Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity research journal*, 24(1):92–96.

Hope Schroeder, Deb Roy, and Jad Kabbara. 2024. Fora: A corpus and framework for the study of facilitated dialogue. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13985–14001.

Dean Keith Simonton. 2012. Taking the us patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity research journal*, 24(2-3):97–106.

Joachim Stempfle and Petra Badke-Schaub. 2002. Thinking in design teams-an analysis of team communication. *Design studies*, 23(5):473–496.

Xiaomin Sun, Yuan Jie, Yilu Wang, Gang Xue, and Yan Liu. 2016. Shared leadership improves team novelty: the mechanism and its boundary condition. *Frontiers in psychology*, 7:1964.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. 2025. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*.

Allison M Traylor, Julie V Dinh, Linnea C Ng, Denise L Reyes, Shannon K Cheng, Natalie C Croitoru, and Eduardo Salas. 2024. It's about the process, not the product: A meta-analytic investigation of team demographic diversity and processes. *Organizational Psychology Review*, 14(3):478–516.

Wei-Chi Tsai, Nai-Wen Chi, Alicia A Grandey, and Sy-Chi Fung. 2012. Positive group affective tone and team creativity: Negative group affective tone and team trust as boundary conditions. *Journal of Organizational Behavior*, 33(5):638–656.

Daan Van Knippenberg, Carsten KW De Dreu, and Astrid C Homan. 2004. Work group diversity and group performance: an integrative model and research agenda. *Journal of applied psychology*, 89(6):1008.

Balázs Vedres and Orsolya Vásárhelyi. 2023. Inclusion unlocks the creative potential of gender diversity in teams. *Scientific Reports*, 13(1):13757.

Jie Wang, Grand H-L Cheng, Tingting Chen, and Kwok Leung. 2019. Team creativity/innovation in culturally diverse teams: A meta-analysis. *Journal of Organizational Behavior*, 40(6):693–708.

Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004):686–688.

Lingfei Wu, Dashun Wang, and James A Evans. 2019. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382.

Gillian B Yeo, Nicole A Celestine, Sharon K Parker, March L To, and Giles Hirst. 2024. A neurocognitive framework of attention and creativity: Maximizing usefulness and novelty via directed and undirected pathways. *Journal of Organizational Behavior*, 45(6):912–934.

Sina Zarrieß, Simeon Junker, Judith Sieker, and Özge Alaçam. 2025. Components of creativity: Language model-based predictors for clustering and switching in verbal fluency. In *Proceedings of the 29th Conference on Computational Natural Language Learning*, pages 216–232.

A Appendix

A.1 Participant Instructions

The following instructions were displayed to participants at the beginning of the study:

Thank you for joining the study. In this study you will first work as a team to come up with ideas to solve certain problems. Then you will be asked to complete a short survey individually.

As a team you will be working on four tasks. In the first three tasks, you will be asked to come up one creative idea for a certain problem. In the last task, you will be asked to come up with unusual uses of a couple common objects. Your answers will be evaluated on their creativity, i.e., they should be both novel and useful.

You can discuss among yourselves and one person from your team needs to submit the most creative answer on behalf of the team within the time limit. You have 12 minutes to complete each task, including discussing among yourselves and submitting the answer. Please be mindful with the time as you will proceed to the next task when the time is up.

Please note that you should NOT use any additional resources (i.e., internet, AI tools, etc.) If you are ready, please click the button below to start your first task.

A.2 Participant and Team Composition

To quantify team composition, we computed team-level diversity indices by aggregating individual attributes within each group. For categorical attributes (e.g., gender, ethnicity, discipline, and employment status), diversity was operationalized using Blau's index, which captures the degree of categorical heterogeneity within a team. For continuous attributes (e.g., age and work experience), diversity was measured as the within-team standard deviation. In all cases, higher diversity values indicate greater heterogeneity among team members, whereas values close to zero indicate more homogeneous team composition. Table 4 summarizes participant characteristics and team-level diversity indices. Figure 3 visualizes the distribution of team compositions across key attributes.

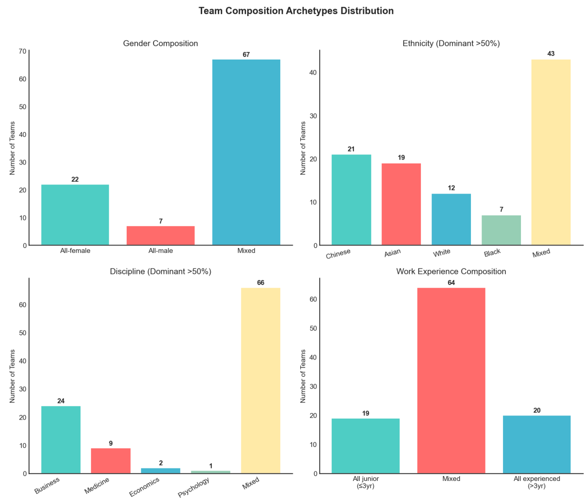


Figure 3: Distribution of team composition across demographic and background attributes.

A.3 Feature Extraction Details

We extracted 18 discourse features operationalizing four theoretically-motivated dimensions from creativity and team research. As described above, consecutive same-speaker utterances were merged into *speaker segments*—units of continuous speech by a single participant—and filtered to retain substantive contributions. Each retained segment was represented using sentence embeddings (all-MiniLM-L6-v2), enabling semantic similarity computation via cosine distance. Table 5 provides a complete list of all features with their factor loadings from subsequent analysis.

Exploration. This dimension captures the breadth and diversity of ideas introduced during discussion. Semantic distance metrics represent one of the most validated computational approaches to creativity measurement, with correlations up to $r = .73$ with human ratings (Beaty and Johnson, 2021). The Divergent Semantic Integration framework (Johnson et al., 2023) demonstrates that average pairwise embedding distance explains substantial variance in creativity judgments. We operationalized exploration through five indicators. *Semantic diversity* was computed as both mean and maximum pairwise cosine distance across segments, capturing the average conceptual spread and the range between the most distant ideas, respectively. To emphasize substantively developed contributions over brief interjections, we also computed *duration-weighted diversity*, where pairwise distances were weighted by segment

Table 4: Overview of participant characteristics and team-level diversity indices with observed ranges.

Category	Key Statistics	Diversity Mean	Diversity Range
Gender	59% Female, 41% Male, <1% Other	0.316	0.00–0.61
Ethnicity	Chinese (27%), South Asian (26%), White (20%)	0.466	0.00–0.78
Employment	54% students; 17% unemployed; 11% full-time	0.401	0.00–0.78
Work exp.	M=4.3 yrs (SD=5.3), 0–40	0.558	0.00–3.05
Disciplines	30 disciplines (top: Business 29%)	0.534	0.00–0.83
Education	55% Masters, 37% Undergraduate	0.339	0.00–0.67
Age	M=27.9 (SD=6.1), 18–60	0.615	0.00–0.72

durations ($w_{ij} \propto \tau_i \tau_j$). *Effective dimensions* captures the dimensionality of the team's semantic space, operationalized as the number of principal components required to explain 90% of embedding variance. Finally, *cluster dispersion* measures whether ideas form distinct thematic clusters versus diffuse spread, computed as the ratio of inter-cluster to intra-cluster distance using k-means clustering.

Convergence. This dimension measures the degree to which teams focus attention toward a shared solution over time. The divergent-convergent model (Guilford, 1967) posits that effective creative processes involve initial exploration followed by narrowing toward selected solutions. Girotra et al. (2010) demonstrated that hybrid processes combining individual ideation with group selection outperform purely interactive brainstorming, highlighting the importance of convergent transitions. We captured convergence through four indicators. *Convergence ratio* measures the relative reduction in semantic diversity from the first to second half of discussion, computed as $(\bar{d}_{\text{early}} - \bar{d}_{\text{late}})/\bar{d}_{\text{early}}$, where positive values indicate progressive focusing. *Diversity trend* captures the temporal trajectory of semantic diversity through linear regression on diversity values computed across sliding windows, where negative slopes indicate convergence. *Final coherence* measures semantic coherence among the final 20% of segments, reflecting solution consolidation. *Final idea alignment* computes the cosine similarity between late-stage segment embeddings and the embedding of the team's submitted solution, measuring whether discussion content converged toward the chosen idea.

Development. This dimension captures elaborative interaction patterns—whether teams build on and extend ideas rather than simply generating disconnected contributions. Information elaboration theory (Van Knippenberg et al., 2004) emphasizes that creativity outcomes depend not merely on idea quantity but on the exchange and integration of information among team members. Coursey et al. (2019) found that elaboration (replies per idea) predicted creativity outcomes more strongly than idea count alone. The SIAM model (Nijstad and Stroebe, 2006) distinguishes flexibility (category transitions) from persistence (within-category exploration), suggesting that both breadth and depth matter for creativity. We operationalized development through four indicators. *Local coherence* measures thematic continuity in idea development through duration-weighted cosine similarity between consecutive segments. *Revisit score* captures whether teams return to and elaborate on initial ideas, computed as the maximum similarity between late-discussion segments and early-discussion content. *Idea deepening ratio* indicates sustained elaboration, operationalized as the proportion of consecutive segment pairs with similarity exceeding a dynamic threshold ($\theta = \bar{c} + 0.5\sigma_c$, where $\bar{c}$ and σ_c are the mean and standard deviation of consecutive similarities). *Semantic momentum* captures consistency in the direction of idea development through cosine similarity between successive direction vectors ($\boldsymbol{\delta}_i = \mathbf{v}_i - \mathbf{v}_{i-1}$).

Participation. This dimension measures how speaking time is distributed across team members. Research on collective intelligence demonstrates that equal distribution of conversational turns is among the strongest predictors of group performance, outperforming individual member ability (Woolley et al., 2010). Nikoleizig et al. (2021) found that speaking time predicted performance evaluations more strongly than actual task competence. We computed participation features on *unfiltered* segments to preserve temporal structure, yielding five indicators. *Speaker entropy* measures equality of speaking time distribution through Shannon entropy normalized by $\log_2 K$ (where K is the number of speakers), with values near 1 indicating equal participation. *Dominant speaker ratio* captures the proportion of total speaking time by the most active speaker. *Segment change rate* reflects interaction pace through the number of speaker alternations per minute. *Response latency* measures discussion fluidity as the mean temporal gap between consecutive segments. *Speaking density* captures the proportion of total discussion time with active speech, distinguishing dense discussions from those with extended silences or pauses.

A.4 Exploratory Factor Analysis Details

Table 5 reports the full factor loading matrix from the exploratory factor analysis (EFA), along with feature retention decisions. Features with primary loadings greater than $|0.50|$ were retained for subsequent analyses. Negative items were reversed to ensure consistent directionality of factor scores.

A.5 Illustrative Discussion Excerpts

This appendix presents excerpts from two discussions on the same task (Voluntary Employee Attrition) that exemplify different process patterns identified through factor analysis.

A.5.1 Example A: High Coherence

Profile: COHERENCE = 2.24 (99th percentile)

High COHERENCE is characterized by sustained elaboration on a single idea thread. In this excerpt, the team identifies employee dissatisfaction as the core problem, then progressively develops a solution (an anonymous feedback platform). Note how speakers reference and synthesize earlier points rather than introducing unrelated ideas. Example transcript excerpts are presented in Table 6.

A.5.2 Example B: High Exploration

Profile: EXPLORATION = 1.65 (100th percentile)

High EXPLORATION is characterized by rapid transitions across semantically diverse topics. In this excerpt, the team generates a wide range of distinct ideas in quick succession: health benefits, salary, career prospects, management, work-life balance, company culture, sabbaticals, and rotation programs. Ideas are introduced but not deeply developed before the next topic emerges. Example transcript excerpts are presented in Table 7.

A.6 Additional Regression Results

Table 8 reports regression results predicting creativity outcomes from discussion process patterns while controlling for task-fixed effects.

Table 8: Process patterns predicting outcomes with task controls.

Pattern	Creativity	Novelty	Usefulness
COHERENCE	$+.007$	$-.008$	$+.028$
EXPLORATION	$+.016^{*}$	$+.032^{*}$	$+.000$
CONVERGENCE	$+.010^{\dagger}$	$+.032^{*}$	$+.004$
PARTICIPATION	$-.003$	$-.011$	$+.013$

$^{*}p < .05$, $^{\dagger}p < .10$. Task fixed effects included.

A.7 Correlation Analyses and Attribute-Level Mechanisms

Table 9 reports pairwise correlations between team composition variables and discussion process patterns. These granular analyses shed light on the mechanisms behind COHERENCE and EXPLORATION.

Social Harmony Facilitates Coherence. While aggregate diversity showed no significant effect on COHERENCE in the regression, *mean Agreeableness* exhibited a strong, positive correlation ($r = .21, p < .001$). This implies that maintaining a coherent, elaborated discussion thread depends more on the interpersonal cooperative tendencies of members than on demographic composition.

Experience Fuels Exploration. Although broad diversity categories showed mixed effects in regression, *Work Experience Mean* ($r = .25, p < .001$) and *Age Mean* ($r = .19, p < .001$) revealed the strongest link with EXPLORATION. This suggests that semantic breadth is primarily fueled by the team's accumulated intellectual capital and life experience rather than simple demographic variety.

Table 5: Exploratory factor analysis: Factor loadings and feature retention decisions.

Dimension	**Feature**	**COHER. (F1)**	**EXPLOR. (F2)**	**CONVERG. (F3)**	**PARTIC. (F4)**	**Decision**
COHERENCE	revisit_score	**.69**	.20	.24	.09	Retain
	local_coherence	**.69**	.05	.02	-.11	Retain
	final_idea_alignment	**.51**	-.25	.18	-.03	Retain
	semantic_diversity_mean	**-.88**	.36	.11	.05	Retain (rev.)
EXPLORATION	effective_dimensions	-.15	**.80**	.20	.21	Retain
	semantic_diversity_max	-.28	**.64**	.11	.09	Retain
	cluster_dispersion	.15	**-.55**	-.26	-.16	Retain (rev.)
	avg_response_latency	.01	**-.67**	.05	.02	Retain (rev.)
CONVERGENCE	convergence_ratio	.09	-.01	**.85**	.04	Retain
	diversity_trend	.01	-.10	**-.82**	-.07	Retain (rev.)
PARTICIPATION	speaker_duration_entropy	-.05	.11	.09	**.75**	Retain
	dominant_speaker_ratio	.14	-.18	-.03	**-.98**	Retain (rev.)
Dropped	semantic_diversity_weighted	-.87	.24	.12	.17	Redundant
	final_coherence	.46	-.35	.28	-.02	Cross-loading
	speaking_density	.30	.53	-.08	-.11	Cross-loading
	segment_change_rate	-.09	.45	-.15	.22	Loading <.50
	idea_deepening_ratio	.26	-.10	-.08	.03	Loading <.50
	semantic_momentum	.18	.05	.02	-.07	Loading <.50

Primary loadings (| >.50) are shown in bold. (rev.) indicates reversed direction for interpretability.
Variance explained: F1=18.3%, F2=15.4%, F3=9.8%, F4=9.5%. Total=53.1%.

Table 6: Example Transcripts for High Coherence

Time	S1	S2	S3
0:34		I mean, I guess the the key thing to maybe try to figure out is why do people generally resign?	
0:41	Yeah, maybe they're not satisfied.		
0:44	So if there was like some sort of program that does like routine.		
0:49			They're not getting rewarded for effort, people.
0:52	To see how like employee employees are finding like how how they see how they are the company like something routinely done say every like 3 months or so, I don't know.		
1:04	So that way you kind of get an idea of where the employee stands on.		
1:09	So they didn't just wake up 1 morning and just decide to quit because they could be like unsatisfied, but like no one knows.		
1:18	Oh, because no one is like bothered to find out how to do not coping with like workload and that kind of thing.		
1:28		So I guess what you're trying to say is there needs to be better monitoring of like employees, sort of.	
1:38		I guess the question for me is like, what prevents people from being able to do that?	
1:46	From just waking up all morning and silent to quit.		
1:49		Yeah, just just feeling unsatisfied and not feeling like you have the ability to communicate to your management that you're unhappy at what?	
2:00			Maybe they are afraid of like possible replications after that.
2:06			Or creative idea to stop voluntary and proliferation?
2:28			So there can be an anonymous channel where employees are asked to showcase their concerns about the job and the management could try to kind of accommodate their requirements.
2:54			Otherwise, write somebody by choice.
3:04		So anonymous platform that allows employees to communicate the grievances or constructive feedback and then having a management system that is very responsive.	
3:20		So it's you said lack, lack of satisfaction, inability for people to communicate it.	
3:25		You were talking about like feeling like you're like undervalued in your role.	
3:29		I think you said that like if it's unappreciated, I guess like salary reimbursement's another key thing like you become unhappy.	
3:38	Upper competitive like salaries.		
3:42	Maybe they seem like company pays like much better than that type of thing.		

Table 7: Example Transcripts for High Exploration

Time	S1	S2	S3
0:30	So health benefits is one that I've been reading about.		
0:36	And in the US, some people are paying more than 10% of their income.		
0:42			Yeah, and I guess like annually, like holiday benefits.
0:45	And also they some insurance, this is crazy.		
0:48	They don't even have access to abortion through the insurance providers that their employers are contracted with.		
0:54	This is like a huge population of people, you know, they're under health insurance that's in place.		
1:00	So that's yeah, health benefits, but then also paternal benefits.		
1:04		Takes like salary or yeah, career prospects or growth or development prospects.	
1:10		Yeah, Management as in lack of belief in management, Yeah.	
1:22		Word life balance, Company culture.	
1:41		It's coming way too easy for us right now.	
1:44		Can you tell I'm trying to leave my company as soon as possible.	
1:46		But yeah, it's literally my life.	
1:51			I think these also like company prospects are like Yep, is it going anywhere?
1:57		Yeah, Yeah, exactly.	
2:04	Also, I think that you just want to gain diverse experiences as well because then you see, I don't know how do I call that one, but you know, you see other people coming in that are new to your company and there was bring some new ideas and you want some of that by gaining those other experiences.		
2:20		You could talk about is it rewarding, is it fulfilling?	
2:25			Oh my God that's literally my 3 criteria for a job.
2:31			Have out of three of those, you know, I'm I can stay, but I should really go.
2:35			But if it's just one out of three, then I will go.
2:38	Rewarding, fulfilling, and and what?		
2:41			It open more doors is the way I have raised in.

Table 9: Correlations between team composition and discussion process patterns.

Variable	COHERENCE	EXPLORATION	CONVERGENCE	PARTICIPATION
Team characteristics				
Team size	$-.18$***	$+.15$***	$-.11$**	$+.21$***
Mean age	–	$+.19$***	–	$+.19$***
Mean work experience	$-.09$**	$+.25$***	–	$+.23$***
Background diversity				
Gender diversity	$-.06$†	–	–	$+.28$***
Ethnicity diversity	–	$+.14$***	–	–
Location diversity	–	$+.18$***	–	–
Education level diversity	–	–	$-.19$***	–
Area of work diversity	$-.12$***	$+.07$*	–	$+.20$***
Personality				
Agreeableness mean	$+.21$***	–	–	$-.09$**
Openness mean	$+.09$**	$+.12$***	–	$+.09$**
Openness diversity	$-.09$**	$-.13$***	–	$-.15$***

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$. – indicates $|r| < .05$.