Zooming in on the Effect of Sociometric Signals on Different Stages of the Design Process

Zooming in on the Effect of Sociometric Signals on Different Stages of the Design Process

Steffi Kohl
Human Data Interaction Lab, Zuyd University of Applied Sciences
Netherlands

Kay Schröder
Human Data Interaction Lab, Zuyd University of Applied Sciences
Netherlands

Mark Graus
Obvion, Maastricht University
Netherlands

Emir Efendic
Maastricht University
Netherlands

Jos G.A.M. Lemmink
Maastricht University
Netherlands

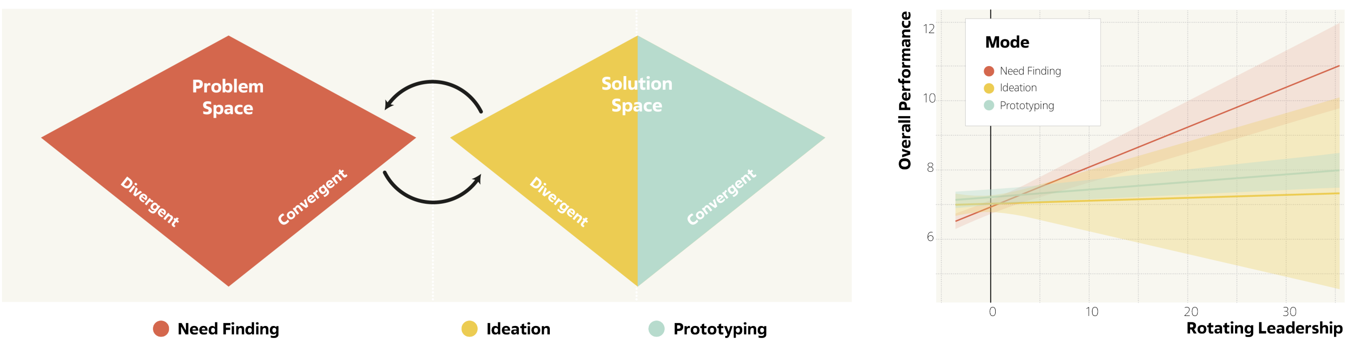

Figure 1: Left: Double diamond framework illustrating how the design thinking modes are separated into convergent and divergent cognitive operations. Right: Interaction plot showing the effects of rotating leadership on team performance for different modes.

ABSTRACT

Collaborative creativity is an essential part of modern teamwork and is often supported by formal techniques, such as design thinking. Current support tools are often limited in scope as understanding the time-varying nature and structure of team communication is insufficient. We investigate how collaborative creative activities in new product development teams can be digitally supported while maintaining face-to-face communication. This work analyzes to what extent paralinguistic and proxemic features of team interaction relate to performance in new product development teams and if and how this relationship differs for different stages in the design process. This is investigated by applying multilevel modeling on data collected during a four-week new product development cycle. The cycle was completed by four teams, during which data were collected automatically using sociometric badges that capture social signals of team interactions. In addition, the data are combined with survey-based measurements on the team's daily design process and periodic performance evaluations. The current paper provides evidence that social signals are related to team performance and that this relationship varies across the stages in the product design process. Certain social signals contribute positively in one stage but less in other stages, showing the importance of using multimodal signals when modeling high-level collaborative patterns. This research contributes to the literature by providing a better understanding of relevant factors when designing supporting tools or methods for collaborative creative problem solving.

C&C '22, June 20–23, 2022, San Servolo, Venice

ACM ISBN 978-1-4503-9327-0/22/06…$15.00
https://doi.org/10.1145/3527927.3532810

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

human-computer interaction, computational paralinguistics, task performance, social signal processing

ACM Reference Format:

Steffi Kohl, Kay Schröder, Mark Graus, Emir Efendic, and Jos G.A.M. Lemmink. 2022. Zooming in on the Effect of Sociometric Signals on Different Stages of the Design Process. In *C&C '22: ACM conference on Creativity & Cognition, June 20–23, 2022, San Servolo, Venice.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3527927.3532810

1 INTRODUCTION

Organizations increasingly realize that teams can be a highly effective working unit when striving for creativity in the workplace

[2]. However, not all teamwork is considered good teamwork and, as Salas and Reyes [31] write, "a team of experts does not necessarily make an expert team." In other words, extensive domain knowledge does not prevent a team from failing if its members do not know how to work together effectively. Indeed, research on collaborative problem solving has found that more often than not, collaboration results in "process loss" instead of "process gain;" group interaction fails to yield performance that exceeds that of the individual group members [12]. Collaborative problem solving requires more than simply joining forces. However, the field remains mute on what determines the success of collaborative problem solving. Teams need to exchange knowledge and information while coordinating skills to stimulate idea generation. We will refer to this process of *communication, coordination and interpretation* as collaborative creative problem solving (CCPS) [13]. While prior research shows that communication contributes to effective CCPS, we possess only limited knowledge of how team communication contributes to performance. Team communication relies on mutual information exchange across various modalities, including verbal (i.e., paralinguistic) and nonverbal (i.e., proxemic) modalities. Paralinguistics relates to all aspects of spoken communication except the semantic content. Proxemics relates to how people use space when communicating. These features have proven to be robust in previous studies of nonverbal communication. They stem from a diverse set of domains such as sociology and psychology [11] and, most recently, human–computer interaction (HCI) [21]. Scholars in these domains have investigated human proxemic and paralinguistic behavior since the 1920s. The utilization of paralinguistic and proxemic modalities has shown to vary across tasks, and some studies have found a structural difference between communication patterns in teams during different CCPS tasks [14, 17].

One reason for the variability across tasks seems to be related to team cognition. Indeed, it has been argued that team communication is an indicator of team cognition as communication reveals cognitive processing at the team level [4, 37]. Further, research has identified that team cognition can vary throughout the CCPS process depending on the group's cognitive state. For example, Stempfle and Badke-Schaub [35] have identified different cognitive operations teams utilized for different tasks during the design thinking process, a method used for CCPS (Figure 1). Consequentially we propose that when investigating communication in teams during CCPS, research should differentiate between tasks when evaluating team performance to account for the underlying cognitive operations utilized by the team members. Throughout this paper, the term *mode* will be used to differentiate these interactions for different tasks within teams. Few studies have investigated the association between different modes and social signals on performance. The current paper is one of the first investigations using proxemics and paralinguistics multimodal models to understand design thinking collaboration dynamics. We aim to demonstrate the value of using these mult-modal signals when modeling high-level collaborative patterns to understand team performance.

This study investigates the relationship between social signals and overall team performance and how that relationship differs across modes. We use a data set collected during real-world meetings of new product development teams to answer this question. Using multilevel modeling, we observe that the relationship between the proxemic feature *rotating leadership* on performance is significantly moderated by mode. We demonstrate that variation between working individually, in small subgroups and in full teams collaborating in close proximity is more beneficial for performance in the need-finding phase of the product development process than in the prototyping phase.

Our contribution is that we have found a measurable interaction effect of *mode* on the relationship of *social signals* on *team performance*. This finding is important as the literature thus far does not differentiate between CCPS tasks, ignoring the underlying cognitive operations of the teams. This highlights a significant shortcoming of the current literature that could limit understanding: the effects of different social signals during specific modes might go unnoticed when they are combined within the same analysis.

We believe that the conclusions drawn from this analysis are applicable to research on team performance and contribute to the foundations for future research aiming to enhance collaborative capabilities, such as through automated support tools for CCPS or novel design thinking process frameworks. Both these research streams need to consider the underlying cognitive operations to help teams perform better.

2 RELATED WORK

2.1 Collaborative Creative Problem Solving Through Design Thinking

Design thinking is commonly used as a methodology by teams working with creative or innovative processes within companies. It involves two or more people engaged in a coordinated attempt to find a joint solution to a problem by establishing common ground that pertains to the problem space and jointly developing a solution that accommodates multiple perspectives [6, 36]. Various definitions of formal methods that underline the design thinking methodology exist in the literature. However, three modes commonly cited within a design thinking approach are need finding (NF), ideation (ID), and prototyping PR. NF is the process of defining the problem. ID is the process of generating ideas and solutions. PR encompasses building models to facilitate the development and selection of concepts [20, 33].

We chose design thinking because it is a commonly applied methodology that entails two very distinguished cognitive operations: divergent and convergent thinking. While divergent thinking aims to find many possible answers or options to a particular problem, convergent thinking narrows down multiple ideas into a single solution [10]. During the design thinking process, teams use convergent and divergent thinking to explore the problem and solution space to apply CCPS successfully. NF inhabits the problem space, utilizing a mix of divergent and convergent thinking. ID and PR share the solution space while utilizing only divergent and convergent thinking, respectively (see Figure 1).

2.2 Modeling Behavioral Patterns During Collaborations

Research on the impact of communication on team performance is spread across various streams of literature. The most prominent

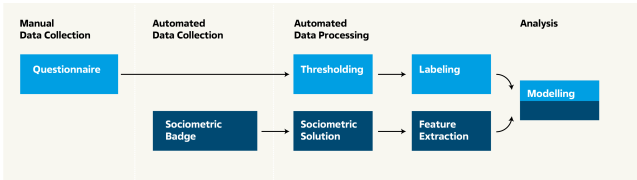

Figure 2: Overview of the data collection and analysis process

research streams are HCI, organizational science, and cognitive science literature. This research has identified a variety of factors found to predict team-level task performance [29]. One prominent research stream focuses on evaluating social signals, such as proxemic and paralinguistic behavior, during team interactions [19, 22, 41]. However, few studies have investigated the association between different modes and social signals on performance. Prior work is often limited to investigating a single type of mode [7, 8, 19, 32], with studies such as Eloy et al. [8] differentiating between goal type but not between modes. Studies that do include multiple modes assume that all will benefit from the same social signals, as seen in the study by Woolley et al. [39] on understanding collective intelligence in teams. While that study asks participants to solve a wide range of tasks, it does not attempt to identify how the relationship between different signals and collective intelligence changes for the different participant modes.

While early work for data-driven modeling on collaboration behavior patterns has mainly aimed to model lower-level behavioral dimensions, such as turn taking [28], recent efforts go beyond low-level signals to model high-level collaborative behavioral patterns. For example, postural markers, such as forward and backward leaning, have been used in human activity recognition to differentiate team member group functions [5]. Proxemic features, such as time spent in close proximity, are indicators of knowledge-sharing dynamics and affect group creativity [9, 15].

As uni-modal features often cannot richly capture complex social interactions, multimodal signals have been increasingly used in modeling high-level collaborative patterns. An example of this is the study by Murray and Oertel [22] that modeled task performance on a team problem-solving task. They trained a random forest classifier to predict task performance from vocal and linguistic features. The multimodal feature set outperformed the uni-modal feature set, demonstrating the added value of multimodal data collection.

Utilizing these signals in successful machine learning models requires theoretical foundations drawn from HCI and organizational science literature with regard to team dynamics. Until recently, these dynamics have remained elusive due to their complexity and lack of quantitative measures. However, wearable electronic devices have made collecting detailed information on team communication affordable. Research indicates that there is predictive power in social signals collected with these devices [27]. This is particularly applicable to collocated collaboration settings because face-to-face teamwork remains the dominant mode for solving complex problems despite the increasing amount of work done by virtual teams. Furthermore, collocated collaboration provides unique benefits that are not easy to achieve in digitally mediated forms of teamwork [26], such as increasing creativity [9] and performance [23]. While preliminary work has demonstrated the feasibility and utility of leveraging multimodal signals to predict behavioral patterns during collaboration activities, more research is needed to understand which data sources best predict certain activities. A detailed overview of the current state of the literature on collocated collaboration modeling using multimodal interaction modeling can be found in Praharaj's [30] comprehensive literature analysis.

3 METHOD

3.1 Data Set

We aim to investigate team-level performance within a collaborative product development task using a data set published earlier [17]. We chose this data set for several reasons. First, it was collected during a design thinking process and provides information on the participants' mode on a given day. Second, the data set contains team-level performance assessed within the context of a real company. This will help us better understand the interplay of social signals and modes with regard to team performance. Lastly, prior analysis of this data set revealed a "sociometric DNA" within the modes expressed as structural differences in the patterns of the individual modes[17]. We believe that understanding how these differences link to performance is the next step toward understanding and improving social signals in team communication.

The data set contains social signals collected from a group of young professionals engaging in a new product development (NPD) sprint exercise at a large consultancy. In total, 18 participants split into four groups of 4–5 members each were observed over 13 days. Team members were unfamiliar with each other before the start of the project. All teams worked without supervision, so each team structured their workdays and scheduled team meetings as necessary. The teams had no formally defined hierarchy, and no formal roles were established. Data on the participants were collected using wearable electronic devices during working hours, excluding lunch breaks. The wearable electronic device used in this data collection is a sociometric badge that automatically measures social signals derived from speech, body motion, and relative location. Sociometric badges are well established in the literature, having been successfully deployed in a variety of organizational contexts with a variety of predictor variables. The badges have been used to predict organizationally relevant outcomes such as job attitude and performance [25], job satisfaction [24], network cohesion [40], creativity [38], group performance [23], and group collaboration [16].

As not all group members were present on all days, only 222 instances of sociometric signals are available out of 234 potential instances. The features used in this study are derived from these speech and proximity signals as provided by the Sociometric Solutions software [34]. All participants were asked to complete daily questionnaires to provide information on the design thinking activities after finishing all work-related activities. We have only included the answers assessing how much time (in percentages) participants spent each day on the three different modes. The response rate to this question was 100%, providing 222 responses. Eight senior consultants assessed the team performance of each group at the location site.

3.2 Variables

Data from several studies suggest that even at small sample sizes, significant patterns can be measured by social signals extracted from the sociometric badges. These signals, hereinafter referred to as features, include turn taking, activity levels, and proximity within the network [27]. Prior work [17] focuses on features that capture the rotation of nonverbal behavior in teams. That research assessed the predictability of design thinking mode using five different rotation features: rotating leadership (RL), rotating contributions (RC), turn taking (TTK), successful interruptions (SI), and unsuccessful interruptions (UI). We used variance inflaction factor (VIF) values to assess if multicollinearity is present to avoid inflated regression coefficients. Due to the multicollinearity of the interruption features (SI & UI) with each other and with turn taking, the interruption features will be excluded from the analysis. We made this decision as turn taking is a well-established feature in the literature [27]. Only the descriptions of RL, RC, and TTK will be covered in the following sections. Furthermore, the dependent variable team performance (Overall Performance) and the moderation variable design thinking mode (MODE) will be explained.

3.2.1 Rotating Leadership. Rotating leadership (RL) in this data set is calculated from the proxemic features measured by the sociometric badges via infrared and Bluetooth signals. It represents the physical location of each participant in relation to the others, and it changes over time. Teams showing high values in RL go through many changes in the number of team members in proximity to each other throughout the day. The groups in this study often split up and worked individually before coming together again to jointly work on the project. The resulting data form a social network evolving through the data collection period. RL reveals changing network structures where people oscillate between peripheral and more central positions. In social network analysis terms, RL is a measure of the frequency in which people change their betweenness centrality in the team when represented as a graph. Betweenness centrality is calculated by dividing the times a node in a network is located on the shortest path by the total number of paths. RL represents the changes within the centrality by counting the number of local maxima and minima over time for a person [17, 18].

This is described in the equation below, where the superscript *BC* indicates that the local minima and maxima are for the betweenness centrality curve, and *i* indicates the person. Higher numbers indicate more rotation of leadership.

$$RL_i = \#localMinima_i^{BC} + \#localMaxima_i^{BC} \tag{1}$$

Research has shown that this measure has a relevant impact on knowledge-sharing dynamics, affects individual and group creativity, and can be a predictor of innovative performance [1, 9]. RL is a day-level variable.

3.2.2 Rotating Contribution. Like RL, rotating contribution (RC) is a consistent indicator of creativity, a key component of CCPS. For both signals, more rotation is positively related to performance during creative tasks, while less rotation is preferable for non-creative tasks [9]. In contrast to RL, which is a proxemic measure of rotation, RC measures paralinguistic rotation. Specifically, it measures the oscillation of the contribution index (CI, or $speaking - listening/speaking + listening$) by counting the number of local maxima and minima in the CI curve of a person over time. It thus represents how frequently people change the amount of time they spend listening vs. speaking [27]. RC is a day-level variable.

3.2.3 Turn Taking. TTK in groups has been associated with collective intelligence and is, in general, considered a measure of the involvement of all team members, which is crucial for team success [3, 39]. Lower numbers of turns taken and increased mean speaking segment length are correlated to group effects such as diminished perceptions of individual and group creativity, as well as lower levels of involvement [27]. Within this data set, turns are speaking segments that occur after and within 10 seconds of another speaking segment. Two speaking segments do not need to be from different people to count as a turn; one person can "self-turn" by pausing and then starting to speak again. This would count as two turns. TTK is a day-level variable.

3.2.4 Team Performance. Overall performance was collected via a survey from senior consultants. The consultants were asked to rate the overall quality of each team's performance at the end of every workweek. The performance measure was assessed on a Likert scale ranging from 1 to 10, with 5 as a neutral point. The team performance measure was measured on a weekly basis and is thus a week-level variable.

3.2.5 Design Thinking Mode. MODE is defined by the percentages reported by the participants. Each observation corresponds to a participant's answer on a given day and is a day-level variable. Each observation was assigned one of four classes: NF, ID, PR or MIXED, depending on whether the participant indicated that at least 60% of the day was spent in the corresponding mode. This cut-off point was chosen to ensure single labeling. For example, an observation indicating 40% NF and 60% ID was assigned to the class ID. The MIXED class was assigned if no mode was present for more than 60% on a given day. MODE is a day-level variable.

4 ANALYSIS AND RESULTS

Multilevel modeling was applied because measurements in the data are not independent. All variables were either: (1) measured per participant (RL, RC, TTK and MODE); or (2) measured per team (overall performance). The participant variables were measured on a daily basis, and the team variable is measured on a weekly basis. To account for our data structure, we conducted mixed effect modeling where individuals were nested in teams, and time was included as a random slope. In the first step, we tested the assumptions and prepared the data. Data inspection revealed that the data is normally distributed, and homogeneity of variance can be assumed. The features were standardized because they were measured at different scales. With the aim of this analysis in mind, we removed all data points assigned the class *MIXED* in MODE, reducing the number of data points from 222 to 156.

4.1 Overall Performance

The guiding research question for this analysis is how different features affect overall performance and how this relationship differs for different modes. To evaluate the moderating effect of MODE, effect coding was used on the remaining three levels of MODE,

Table 1: Log Likelihood Test

Model	AIC	LogL	Chisq	Df	Pr(>Chisq)
Unconditional Means	381.36	−186.68			
Random Slope	215.43	−99.71	173.94	4	0.0000
Random Slope	215.43	−99.71			
Perf - RL*MODE	203.71	−88.86	21.71	5	0.0006
Perf - RC*MODE	212.81	−93.41	12.62	5	0.0273
Perf - TTK*MODE	210.55	−92.27	14.88	5	0.0109

Note: Akaike information criterion (AIC); Log likelihood values (LogL) χ^2 (Chisq); Degrees of freedom (df)

creating two regressors. We used the class *PR* in MODE as a contrasting group, so observations in this class were assigned a −1. Thus, we created one regressor named *NF-PR* for modes of the NF class assigned a 1 and another named *ID-PR* for observations of the ID class assigned a 0. In line with the recommended procedure for multilevel analysis, we developed a sequence of models from the simplest to the most complex. We compared model fit using AIC, which penalizes for model complexity. If the AIC is lower for a more complex model, then the gain in fit is worth the extra complexity. In addition, we also conducted likelihood ratio tests for all presented models to assess if the model fit improvement is statistically significant. All models use overall performance as the dependent variable.

We start by building the most basic model, an unconditional means model. This model has no predictors and allows the intercept to vary for participants nested in teams. This model lets participants have their own baseline values but assumes that participants respond to time in exactly the same way. The next step in model fitting is to build a random intercept and slope model. In addition to allowing the intercept to vary for participants nested in teams, the random slope model allows each group line to have a different slope, hence allowing the explanatory variable to have a different effect for each group over time. We conduct a likelihood ratio test using maximum likelihood estimates to see if allowing the slope to vary per week improves our model fit. The results in Table 1 show a significant improvement from the unconditional mean model to the random slope model, increasing χ^2 by 173.94, ($df = 4, p < 2.2e-16$). The improved fit is also indicated by the AIC score, which is lower for the random slope model. As this model not only shows a better model fit but also makes sense for our data structure, we will be using a random slope model structure when adding our level-one predictors.

Table 1 shows the results of the likelihood ratio test. Six different models were generated, but only models with significant coefficients for the predictors are presented in the tables. Adding the predictors by themselves did not yield significant coefficients. However, all moderation models showed significant coefficients. Evaluating the model fit between all models, we can observe that Model *Perf - RL*MODE* (Table 1) outperforms all models with an AIC score of 203.71. Adding the predictors MODE and RL to this model provides the largest χ^2 improvements. Table 2 shows the results for the random slope models with the interaction effects

Table 2: Mixed Effect Models for Overall Performance

	Overall Performance		
	RL*MODE	RC*MODE	TTK*MODE
RL	0.049^{**}		
	(0.017)		
RC		−0.001	
		(0.008)	
TTK			0.00004
			(0.00003)
NF-PR	-0.127^{*}	-0.176^{**}	-0.183^{**}
	(0.059)	(0.058)	(0.059)
ID-PR	−0.031	0.018	0.011
	(0.053)	(0.054)	(0.055)
RL:NF-PR	0.067^{*}		
	(0.027)		
RL:ID-PR	-0.040^{*}		
	(0.018)		
RC:NF-PR		−0.004	
		(0.013)	
RC:ID-PR		0.010	
		(0.010)	
TTK:NF-PR			−0.00001
			(0.00003)
TTK:ID-PR			0.00002
			(0.00003)
Constant	7.058^{***}	7.019^{***}	7.027^{***}
	(0.101)	(0.107)	(0.114)
Observations	156	156	156
Log Likelihood	−104.595	−110.350	−126.408
Akaike Inf. Crit.	235.190	246.701	278.816
Bayesian Inf. Crit.	274.839	286.349	318.464

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

between the factors and MODE. Looking at the models, we can observe a negative direct effect of the regressor NF-PR, which indicates that the NF class of MODE contributes negatively to overall performance compared to the contrasting group PR, which was coded with −1. There seems to be no significant effect of the class ID on performance compared to PR as measured by the regressor ID-PR. Looking at the interaction effect specifically, we can observe that the relationship between RL and performance is significantly moderated by MODE. The expected change in the overall performance for a one-unit increase in RL when a team is in the NF class of MODE is 0.067 higher than when a team is in the PR class. The expected change in the overall performance for a one-unit increase in RL when a team is in the ID class of MODE is −0.04 lower than a team in the PR class (Table 2 – Model *RL*MODE*). This is to say, for NL, an increase in RL is related to an increase in overall performance. This effect is not as strong (PR) or even not detected (ID) for the other modes. These relationships are visualized in the interaction plot Figure 1. We do not observe a significant moderation effect for RC or TTK (Table 2 – Model *RC*MODE* & *TTK*MODE*)

5 DISCUSSION AND CONCLUSION

Following our analysis, we can answer our research question. Our results show differences in the effect of social signals on team performance for different participant modes. Different features (specifically RL) affect overall team performance, and this relationship differs over different design thinking modes. While all modes benefit from increased RL, NF does so significantly more than ID and PR. More research is necessary to understand why we observe this relationship. A likely explanation is that while high RL values are associated with positive performance in creative tasks, the directionality of the correlation changes for non-creative activities. The wide spread of variance within the ideation mode and the overall lower impact on the performance mode might indicate that these modes include more non-creative activities as they occur later in the design thinking process.

The key contribution of our paper is the investigation into the relationship between proxemic and paralinguistic features and performance for different modes, which until now has not been explored to a large extent. By analyzing a data set of fairly limited rotational features, this research demonstrates that communication patterns have correlations with team performance and that these correlations differ across different modes. These findings can guide future research designs aimed at understanding the relationship between communication and team performance.

Further, we are adding to the limited knowledge pool of understanding the relationship between social signals and team performance. This contribution to the existing knowledge is vital as it can help to analyze team interaction dynamics during CCPS. Most work on team performance within the social signal processing community has focused on single-mode analysis. The impact of mode on performance has been underexplored. We demonstrate the usefulness and necessity of differentiating between modes when understanding proxemic and paralinguistic features in relation to task performance by utilizing those features captured in a real-world work environment.

Future work will look at different features and how to combine them to yield further improvements. In addition, more modes should be explored to understand how paralinguistic and proxemic features relate to different types of CCPS.

Furthermore, the proposed relationship should also be evaluated in a more controlled setting to examine the effects with less noise. Such studies should separately evaluate the modes without the interaction between them that is observed in the current study. Such examinations could also include a more focused performance evaluation that evaluates the performance of the individual mode instead of overall performance.

Finally, these findings pave the way for the exploration of automated coaching of teams by creating interventions to guide teams to interact in ways that should result in optimal team performance for their current mode.

REFERENCES

[1] Fernando G Alberti, Federica Belfanti, and Jessica D Giusti. 2021. Knowledge exchange and innovation in clusters: a dynamic social network analysis. *Industry and Innovation* 28, 7 (2021), 880–901.

[2] Neil Anderson, Kristina Potočnik, and Jing Zhou. 2014. Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of management* 40, 5 (2014), 1297–1333.

[3] Maria Chuy, Jianwei Zhang, Monica Resendes, M Scardamalia, and C Bereiter. 2011. Does contributing to a knowledge building dialogue lead to individual advancement of knowledge. In *Connecting computer-supported collaborative learning to policy and practice: CSCL2011 conference proceedings*, Vol. 1. ISLS, 57–63.

[4] Nancy J Cooke, Andrew Duchon, Jamie C Gorman, Joann Keyton, and Anne Miller. 2012. Preface to the special section on methods for the analysis of communication. *Human Factors* 54, 4 (2012), 485–488.

[5] Julius Dietzel, R Elena Francu, Benjamin Lucas, and Mohamed Zaki. 2018. Contextually Defined Postural Markers Reveal Who's in Charge: Evidence from Small Teams Collected with Wearable Sensors.

[6] Kees Dorst. 2011. The core of 'design thinking'and its application. *Design studies* 32, 6 (2011), 521–532.

[7] Sidney D'Mello, Angela EB Stewart, Mary Jean Amon, Chen Sun, Nicholas Duran, and Valerie Shute. 2019. Towards Dynamic Intelligent Support for Collaborative Problem Solving. In *CEUR Workshop Proceedings*, Vol. 2501. CEUR-WS, 59–65.

[8] Lucca Eloy, Angela EB Stewart, Mary Jean Amon, Caroline Reinhardt, Amanda Michaels, Chen Sun, Valerie Shute, Nicholas D Duran, and Sidney D'Mello. 2019. Modeling Team-level Multimodal Dynamics during Multiparty Collaboration. In *2019 International Conference on Multimodal Interaction*. 244–258.

[9] Peter A Gloor, Adam Almozlino, Orr Inbar, Wei Lo, and Shannon Provost. 2014. Measuring team creativity through longitudinal social signals. *arXiv preprint arXiv:1407.0440* (2014).

[10] J. P. Guilford. 1950. Creativity. *American Psychologist* 5, 9 (1950), 444–454. https://doi.org/10.1037/h0063487

[11] Jinni A Harrigan. 2005. Proxemics, kinesics, and gaze. *The new handbook of methods in nonverbal behavior research* (2005), 137–198.

[12] Friedrich Hesse, Esther Care, Juergen Buder, Kai Sassenberg, and Patrick Griffin. 2015. A framework for teachable collaborative problem solving skills. In *Assessment and teaching of 21st century skills*. Springer, 37–56.

[13] Otmar Hilliges, Lucia Terrenghi, Sebastian Boring, David Kim, Hendrik Richter, and Andreas Butz. 2007. Designing for collaborative creative problem solving. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*. 137–146.

[14] Dinesh Babu Jayagopi, Taemie Kim, Alex Sandy Pentland, and Daniel Gatica-Perez. 2010. Recognizing conversational context in group interaction using privacy-sensitive mobile sensors. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 8.

[15] Yared H Kidane and Peter A Gloor. 2007. Correlating temporal communication patterns of the Eclipse open source community with performance and creativity. *Computational and mathematical organization theory* 13, 1 (2007), 17–27.

[16] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: enhancing group collaborationusing sociometric feedback. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 457–466.

[17] Steffi Kohl, Mark P Graus, and Jos GAM Lemmink. 2020. Deciphering the Code: Evidence for a Sociometric DNA in Design Thinking Meetings. In *International Conference on Human-Computer Interaction*. Springer, 53–61.

[18] Steffi Kohl, Mark P Graus, and Jos GAM Lemmink. 2022. Context is key: Mining social signals for automatic task detection in design thinking meetings. In *International Conference on Human-Computer Interaction*. Springer.

[19] Uliyana Kubasova, Gabriel Murray, and McKenzie Braley. 2019. Analyzing verbal and nonverbal features for predicting group performance. *arXiv preprint arXiv:1907.01369* (2019).

[20] Jeanne Liedtka. 2015. Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. *Journal of product innovation management* 32, 6 (2015), 925–938.

[21] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 331–338.

[22] Gabriel Murray and Catharine Oertel. 2018. Predicting group performance in task-based interaction. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 14–20.

[23] Daniel Olguın Olguın, Peter A Gloor, and Alex Sandy Pentland. 2009. Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS*, Vol. 9.

[24] Daniel Olguín Olguín, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. 2008. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 1 (2008), 43–55.

[25] Daniel Olguin Olguin. 2011. *Sensor-based organizational design and engineering*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[26] Judith S Olson, Stephanie Teasley, Lisa Covi, and Gary Olson. 2002. The (currently) unique advantages of collocated work. *Distributed work* (2002), 113–135.
[27] John N Parker, Edgar Cardenas, Alexander N Dorr, and Edward J Hackett. 2018. Using sociometers to advance small group research. *Sociological Methods & Research* (2018), 0049124118769091.
[28] Alex Pentland. 2010. *Honest signals: how they shape our world.* MIT press.
[29] Anna Penzkofer, Philipp Müller, Felix Bühler, Sven Mayer, and Andreas Bulling. 2021. ConAn: A Usable Tool for Multimodal Conversation Analysis. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 341–351.
[30] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus M Specht. 2021. Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning AnalyticsCan We Go the Whole Nine Yards. *IEEE Transactions on Learning Technologies* (2021).
[31] Eduardo Salas, Denise L Reyes, and Susan H McDaniel. 2018. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* 73, 4 (2018), 593.
[32] Kay Schröder and Steffi Kohl. 2022. ViCon - Towards understanding visual support systems in collaborative video conferencing. In *International Conference on Human-Computer Interaction*. Springer.
[33] Victor P Seidel and Sebastian K Fixson. 2013. Adopting design thinking in novice multidisciplinary teams: The application and limits of design methods and reflexive practices. *Journal of Product Innovation Management* 30 (2013), 19–33.
[34] Sociometric Solutions. 2014. Sociometric badge 03-02: Preliminary user guide. *Boston, MA: Author* (2014).
[35] Joachim Stempfle and Petra Badke-Schaub. 2002. Thinking in design teams-an analysis of team communication. *Design studies* 23, 5 (2002), 473–496.
[36] Angela EB Stewart, Hana Vrzakova, Chen Sun, Jade Yonehiro, Cathlyn Adele Stone, Nicholas D Duran, Valerie Shute, and Sidney K D'Mello. 2019. I Say, You Say, We Say: Using Spoken Language to Model Socio-Cognitive Processes during Computer-Supported Collaborative Problem Solving. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.
[37] Dominique Engome Tchupo, Jung Hyup Kim, and Gretchen A Macht. 2020. Fuzzy cognitive maps (FCMs) for the analysis of team communication. *Applied ergonomics* 83 (2020), 102979.
[38] Priyamvada Tripathi and Winslow Burleson. 2012. Predicting creativity in the wild: Experience sample and sociometric modeling of teams. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1203–1212.
[39] 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 (2010), 686–688.
[40] Lynn Wu, Benjamin N Waber, Sinan Aral, Erik Brynjolfsson, and Alex Pentland. 2008. Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an it configuration task. *Available at SSRN 1130251* (2008).
[41] Shun-Chang Zhong, Bo-Hao Su, Wei Huang, Yi-Ching Liu, and Chi-Chun Lee. 2020. Predicting Collaborative Task Performance using Graph Interlocutor Acoustic Network in Small Group Interaction. *Proc. Interspeech 2020* (2020), 3122–3126.