Transdisciplinarity and the Future of Engineering
B.R. Moser et al. (Eds.)

doi:10.3233/ATDE220655

Measuring Characteristics and Influence of Fluctuating Teamwork Processes Based on Natural Language Processing: The Relationship Between Equal Participation and Creativity

Sixiong PENG and Katsuya TORII
University of Tokyo, Japan

Abstract. To leverage different skills of transdisciplinary teams, understanding team processes critical for team performance is crucial. This study examined temporal participation equality and creativity using behavioral data and natural language processing. Twenty-one teams of four people generated business ideas for post COVID19 societies in teams. We segmented their discussion based on lexical similarity and clarified the relationship between each segment and generated ideas. Then we calculated the equality of turn-taking and speaking time and examined their relationship with creativity scores of ideas using hierarchical regression analysis. The result suggested temporal participation equality did not relate to creativity. We discussed its implication and future studies.

Keywords. Teamwork, natural language processing, equal participation, creativity

Introduction

Modern organizations utilize transdisciplinary teams to address complex socio-technical issues. Although transdisciplinary teams can leverage different skills, knowledge, ideas, and perspectives of diverse team members, they are more prone to collapse than homogenous teams due to different mental models of team members [1]. Investigating teamwork that critically influences team performance is needed for supporting transdisciplinary teams.

Traditional research for teamwork mainly relies on surveys and qualitative observation. Although this research has yielded significant progress, these methods have recently been criticized for several reasons [2]. The accuracy of the survey method depends on the respondents' ability of memory retrieval and is prone to the tendency to remember the peak moment and the last moment [3]. It also treats teamwork as stationary, which is true in only limited cases [4][5]. Qualitative observation is also prone to subjectivity and is hard to be applied to a large-scale study. In response to this criticism, an increasing number of studies have adopted behavioral data such as chat logs, work logs, audiovisual records, and physiological data [6]. These measurements are objective and suitable for capturing changes in team states over time. In addition, because it is an unobtrusive measure, it can be utilized for real-time teamwork support tools. This advantage is significant given that data acquisition has become increasingly easy in recent years due to the diffusion of remote work and advances in sensor technology [7].

Drawing on these advantages, this study utilized behavioral data to analyze the relationship between participation equality and creativity. Scholars argued that participation equality influenced team performance [8][9] although empirical results were mixed [10]. Most of these studies examined participation equality in the whole period of teamwork, which ignored fluctuation of temporal participation level. Essentially, overall participation equality and temporal participation equality capture different phenomena. The former reflected on more stable interpersonal relationships of team members such as expectation states [11] or psychological safety [12], whereas the latter reflected on interaction patterns influenced by a local context such as discussion topics and emotion. Few studies pointing out the relationship between temporal participation level and creativity were based on qualitative observation [13][14], calling for more quantitative approaches. Another challenge is an examination of causality. The socio-cognitive framework of team innovation suggested that team processes influence team outcomes through individual or team cognitive processes [8]. Cognitive processes are partly manifested in discussion contents in teamwork. Thus, if participation equality can be connected to discussion contents that are related to outcome creativity, more reliable evidence of causality could be inferred. We attempt to achieve this using the natural language processing technique. The aim of this paper is twofold: testing the relationship between participation equality and creativity, and showcasing analysis of behavioral data using natural language processing and revealing its advantages and limitations.

1. Methods

1.1. Participants

This study analyzed the workshop "Business ideas for post-COVID-19 society." Eighty-four participants were publicly recruited via Social Networking Services and grouped into 21 four-person teams. Their ages ranged from 19 to 33 (mean = 23.1, SD = 3.0). Twelve of them were working adults and the rest of them were university students. Thirty-nine of them identified themselves as female. All members are new to each other. These temporary newly formed teams are similar to teams in event-based workshops in companies.

1.2. Tasks and data

We asked participants to create business ideas for the post-COVID19 society in teams through a process based on design thinking [15]. Before the workshop date, participants answered the questionnaires about their personality including creative efficacy [16] and basic personal information such as affiliation and sex. Participants joined the workshop online using Zoom. Also, they used Apisnote (https://www.apisnote.com/) [17] for recording and organizing their discussion results during group work. The workshop started with an introduction of the workshop theme and a task process overview from a facilitator. Before the group work started, participants individually created ideas (task 0), which were not used in this study. Then group work started with a self-introduction for 15 minutes (task 1). After this, teams were given 90 minutes to share their knowledge about the social changes caused by COVID-19 and analyze this information for generating novel needs (task 2). These needs were essentially pre-inventive structures that served as seeds for subsequent idea generation. Then they individually generate as many business ideas as possible in the next 20 minutes (task 3). Finally, they shared their ideas with team members and chose the best one as a team in the final 60 minutes (task 4). This study only used an audio recording and its transcription for task 2 and textual information of ideas on Apisnote boards in task 3.

1.3. Idea evaluation

We evaluated ideas in task 2 following the Consensual Assessment Technique [18], which is based on the agreement of domain experts about what is creative. Although it requires finding domain experts who internalize domain standards through education and experience, this was difficult to achieve in our context, which included a large number of business ideas in post-COVID19 society. Thus, we recruited three university students who studied methods for creating innovative ideas. Several studies showed that evaluations by students as semi-experts sere as reliable as those by experts depending on domain characteristics [19] [20]. Two evaluators evaluated each idea for its novelty and its effectiveness with a score ranging from 1 to 5 and summed them to calculate a creativity score. We excluded ideas that were too vague or could not be understood as an idea such as needs. We also excluded ideas that were identical to the idea generated in task 0 because these were not related to the group work. As a result, we evaluated 475 ideas in total. Inter-rater agreement was low (Cronbach's alpha = 0.41). The main reason for the low agreement is the limited information used for evaluation. In our workshop, participants spent only 20 minutes thinking of multiple ideas, which did not allow them to fully elaborate their ideas. Hence, many ideas left room for interpretation. Discussion with evaluators revealed that they often supplemented information during evaluation for making ideas clear enough to evaluate and this mental process could be the source of different evaluations. Thus, we decided to re-evaluate ideas that had more than a 1-point difference in scores between evaluators. The second evaluation was conducted after asking evaluators to write down reasons for their first evaluations and sharing this information among all evaluators. This second evaluation yielded higher inter-rater agreement (Cronbach's alpha=0.71). We used the second evaluation for subsequent analysis.

1.4. Analytical procedure

Figure 1 summarizes our analytical procedure. We firstly segmented discussion based on the Text Tiling algorithm [21], which detects the points of subtopic changes in a document by comparing the lexical similarity between segments before and after every possible point. We made some modifications for applying it to the conversation. We chose the end time of every speaking turn as a possible point for the topic change. Also, we grouped k speaking turn that contained more than 4 letters (in Japanese) as a block to calculate lexical similarity. For lexical similarity, we use cosine similarity of vectors composed by tf-idf scores of every word. Figure 2 illustrates the specific procedure of the modified version of the Text Tiling algorithm. As for parameters, we set k=6, s=2, and n=1 (s and n are the same parameters defined in [21]). We chose these parameters to identify relatively smaller topic changes.

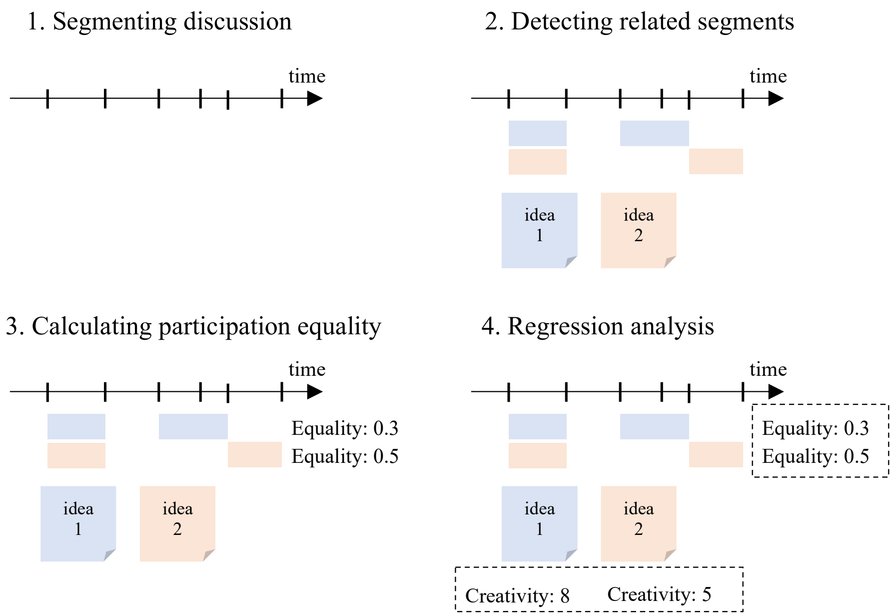

Figure 1. Overview of the analytical process

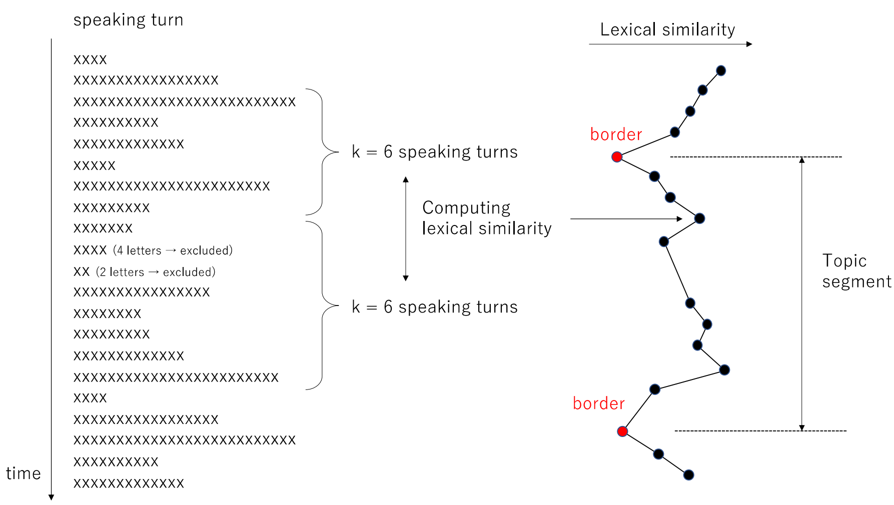

Figure 2. The modified version of the Text Tiling algorithm

Because we would judge whether a topic segment was related to any of the ideas, we expected that one topic segment did not contain more than 1 topic. After inspecting some topic segments, we found topic segments of two teams frequently contained more than 1 topic. These teams spent most of their time sharing thoughts in turn without elaborating on them, resulting in topic change at nearly every speaking turn. We compared lexical similarity between six speaking turns and detected borders at a minimal similarity. Hence, if topics changed for every speaking turn, the similarity would not be minimal at the point of the topic change. We excluded these 2 teams for subsequent analysis.

In the second step, we calculated lexical similarity between every combination of ideas and topic segments. As for information on ideas, we used sentences written on the Apisnote board. We used the same method for calculating lexical similarity as one that we used for segmenting discussion. Figure 3 illustrates the lexical similarity between one idea and every topic segment arranged by time. The x-axis is time and the y-axis is lexical similarity. It demonstrates that discussion in a topic segment of around 130 minutes is more lexically similar to the idea than in other segments. After checking several graphs, we decided to choose topic segments with lexical similarity above 0.05 as candidates for related segments.

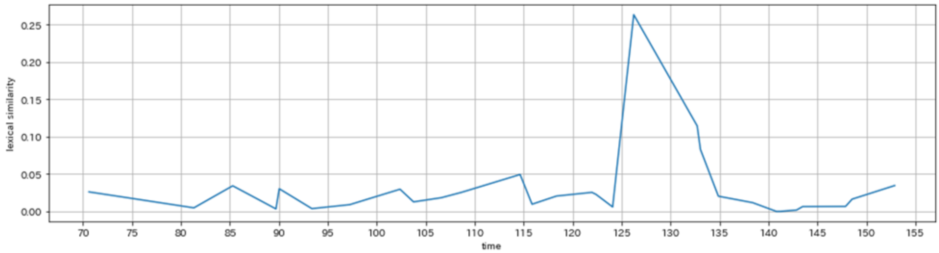

Figure 3. The lexical similarity of an idea and discussion segments

After inspecting several related segments, we found the accuracy of this method questionable. There were false-positive cases where discussion segments that were not related to ideas were detected. This case happened when several words used in ideas were used in discussion under unrelated contexts. Also, there were false-negative cases where some keywords in the discussion were paraphrased in an idea. Although we could improve the accuracy by considering different methods for calculating lexical similarity, we left this as a future study and manually excluded false-positive cases for this study. Out of 1210 segments that were related to any of the ideas, 589 segments were accurate. The high proportion of false-positive cases was partially due to our intentionally low threshold. Although the mean of lexical similarity of these segments was significantly higher than false-positive segments ($t=14.51$, $p<0.00$), the number of overlapping cases was not trivial, indicating difficulty to improve accuracy by simply changing the threshold.

Next, we calculated participation equality in all the related segments for each idea. Based on previous research [22], we calculated the coefficient of variation of the number of turn-takings and the coefficient of variation of the speaking time. A lower value indicated more equal participation.

Lastly, we tested the relationship between participation equality and the creativity score. Because participation equality of discussion in the same team was not independent, we adopted a hierarchical linear model. We controlled average creative efficacy measured by questions of [16], the gender balance measured by the number of female members, and the age diversity measured by the standard deviation of members' age considering these effects on creativity.

2. Result

Table 1 and Table 2 illustrate the result of hierarchical regression analysis with participation equality calculated by turn-taking and speaking time, respectively. In the first model, we put control variables centered among teams and participation equality as the level 1 variable that was centered within a group. In the second model, we put the interaction effect between control variables and participation equality to check whether the relationship between participation equality and creativity changed depending on the control variables. In the third model, we put mean participation equality as the level 2 variable to compare the effect of overall participation equality on creativity and that of temporal participation equality.

The results show that level 1 participation equality was not correlated with the creativity score of related ideas in any model. The interaction effect between level 1 participation equality measured by turn-taking and average creative efficacy was weakly significant, which means level 1 participation equality measured by turn-taking may have a relationship with creativity score if teams had high average creative efficacy. Also, level 2 participation equality had a weakly significant relationship correlation with creativity.

Table 1. Hierarchical regression model between participation equality (coefficient of variation of turn-taking) and creativity score

Variables	Model 1	Model 2	Model 3
Intercept	5.75**	5.77**	5.75**
Level 1 Participation equality (turn-taking)	0.03	0.07	0.08
Average creative efficacy	0.47**	0.50**	0.61**
Gender balance	0.09	0.10	0.12
Age diversity	0.06	0.06	0.09†
Level 1 Participation equality (turn-taking) * Average creative efficacy		-0.63†	-0.71
Level 1 Participation equality (turn-taking) * Gender balance		-0.04	-0.05
Level 1 Participation equality (turn-taking) * Age diversity		0.07	0.06
Level 2 Participation equality (turn-taking)			-0.59†
Level 1 Participation equality (turn-taking) * Level 2 Participation equality (turn-taking)			0.32
AIC	1007.3	1003.2	1006.6
BIC	1030.1	1037.4	1059.9
Deviance	985.17	981.70	978.63

Notes: †$p<0.1$, *$p<0.05$, **$p<0.01$

Table 2. Hierarchical regression between participation equality (coefficient of variation of speaking time) and creativity score

Variables	Model 1	Model 2	Model 3
Intercept	5.75**	5.75**	5.74**
Level 1 Participation equality (speaking time)	-0.31	-0.33	-0.31
Average creative efficacy	0.50	0.50**	0.56**
Gender balance	0.09	0.10	0.09
Age diversity	0.06	0.06	0.08
Level 1 Participation equality (speaking time) * Average creative efficacy		-0.15	-0.21
Level 1 Participation equality (speaking time) * Gender balance		-0.10	-0.09
Level 1 Participation equality (speaking time) * Age diversity		0.13	0.12
Level 2 Participation equality (speaking time)			-1.04†
Level 1 Participation equality (speaking time) * Level 2 Participation equality (speaking time)			1.90
AIC	1001.7	1007.1	1007.2
BIC	1036.0	1052.8	1060.6
Deviance	983.73	983.14	979.24

Notes: $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$

3. Discussion

The result indicated that participation equality measured in the related discussion did not relate to creativity while overall participation equality might relate to creativity. This result could be explained by a confounding factor influencing overall participation equality and creativity. Previous research demonstrated that participation equality affects team performance via promoting sharing and elaborating knowledge [9] [23]. In the context of creativity, sharing and elaborating unique knowledge leads to new combinations of knowledge or breaking the widely-held bias that would not happen in individual work [9] [24]. We could not find this effect but found another possible explanation that a confounding factor influenced both participation equality and creativity, which meant that participation equality did not have a causal relationship with creativity. In other words, a creative team was creative because it had some traits that made members participate equally and not because they participated equally. This view is consistent with several studies that utilize teamwork visualizing tools to promote equal participation. Both [25] and [26] realized more equal participation through their tools, but they could not find the relationship between participation equality and team performance. The confounding factor can be relatively stable team emergent states such as psychological safety or cohesion considering temporal participation equality does not correlate with creativity.

Another explanation of our result was the failure to adequately detect related discussions. Although Text Tiling was reported to match with human segmentation of subtopics in documents, its accuracy to segment conversation may not be reliable. Indeed, we found many cases with inadequate segmentation in 2 teams that we excluded as explained in 1.3. Some segmentation of the rest 19 teams might be inadequate as well. In addition, even if segmentations were accurate, the detection of relating segments to ideas could be inaccurate as we discussed in the previous section about the research method. Especially, we did not check the false-negative case. A theory of creative thinking suggested that a combination of unrelated topics leads to a creative idea [27]. In this view, the generated idea may not be lexically similar to the source knowledge. A more adequate method that can capture lexically dissimilar but related segments is needed.

Although our result was not as we expected, this research had several contributions to the research of teamwork in interdisciplinary teams. Firstly, our result suggested that the relationship between participation equality and creativity might be explained by confounding factors such as psychological safety or cohesion rather than a causal relationship. Future studies can consider measuring variables that are expected to correlate with both participation equality and creativity for a more proper understanding of the relationship. Additionally, our analytical process that detected discussion of interest using natural language processing can be applied to different contexts. If one is interested in a discussion that affects any outputs of interest, one can utilize our process although our process of segmenting conversation and detecting related segments should be improved to provide a reliable result.

Our study had several limitations. First, our sample size was low. Our result based on the teamwork of 21 teams may not be generalizable. In addition, our teams were temporary newly-formed virtual teams. Hence our findings may not be generalized to other situations such as existing teams or in-person team discussions. Furthermore, there could be problems with our measurements. Although we carefully revised the evaluations on the creativity of ideas, the evaluation could still not be reliable mainly because of a little amount of information about ideas. Evaluators' interpretation of ideas could be different from creators' understanding of ideas. Similarly, the accuracy of topic segmentation and their relationship with ideas could be questionable as we discussed above.

4. Conclusion

Drawing on the recent trend to utilize behavioral data for analyzing teamwork, this study examined the relationship between participation equality and creativity and showcased the analysis of behavioral data using natural language processing. Our result revealed that participation equality in discussion related to an idea did not correlate with its creativity while participation equality throughout teamwork had a weakly significant correlation with creativity. This result could be possibly explained by confounding factors such as team emergent states. We also demonstrated one method of analyzing temporal teamwork using natural language processing. We hope our study will stimulate future studies of transdisciplinary teamwork.

Acknowledgment

We thank Kazuo Hiekata for supporting workshops and providing constructive feedback.

References

[1] S. K. Horwitz, and I. B. Horwitz, The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography, *Journal of Management*, 2007, Vol. 33, No. 6, pp. 987–1015.
[2] S. W. J. Kozlowski and G. T. Chao, “Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods,” *American Psychologist*, 2018, vol. 73, no. 4, pp. 576–592.
[3] D. Kahneman, B. L. Fredrickson, C. A. Schreiber, and D. A. Redelmeier, “When More Pain Is Preferred to Less: Adding a Better End,” *Psychological Science*, 1993, vol. 4, no. 6, pp. 401–405.
[4] P. J. Ramos-Villagrasa, P. Marques-Quinteiro, J. Navarro, and R. Rico, “Teams as Complex Adaptive Systems: Reviewing 17 Years of Research,” *Small Group Research*, 2018, vol. 49, no. 2, pp. 135–176.
[5] S. Peng, T. Amakasu, H. Kawauchi, H. Horii, and K. Hiekata, Development of Method for Visualizing Behavioral States of Teams, *Advances in Transdisciplinary Engineering*, 2021, Vol. 16, pp. 537-546.
[6] S. Davila-Montero, J. A. Dana-Le, G. Bente, A. T. Hall, and A. J. Mason, Review and Challenges of Technologies for Real-Time Human Behavior Monitoring, *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 1, pp. 2–28, 2021.
[7] K.M. Kniffin et al., COVID-19 and the workplace: Implications, issues, and insights for future research and action, *American Psychologist*, 2021, vol. 76, no. 1, pp. 63–77.
[8] S.B.F. Paletz and C.D. Schunn, A Social-Cognitive Framework of Multidisciplinary Team Innovation, *Topics in Cognitive Science*, 2010, vol. 2, no. 1, pp. 73–95.
[9] C.K.W. De Dreu and M.A. West, Minority dissent and team innovation: The importance of participation in decision making., *Journal of Applied Psychology*, 2001, vol. 86, no. 6, pp. 1191–1201.
[10] J.A. Bonito and A.B. Hollingshead, Participation in small groups, *Annals of the International Communication Association*, 1997, vol. 20, no. 1, pp. 227–261.
[11] J. Berger, B.P. Cohen, and M. Zelditch, Status Characteristics and Social Interaction, *American Sociological Review*, 1972, vol. 37, no. 3, pp. 241–255.
[12] A. Edmondson, Psychological Safety and Learning Behavior in Work Teams, *Administrative Science Quarterly*, Jun. 1999, vol. 44, no. 2, pp. 350–383.
[13] K. Sawyer, *Group genius: The creative power of collaboration*. Basic books, New York, 2008.
[14] L. A. van Oortmerssen, C. M. J. van Woerkum, and N. Aarts, When Interaction Flows: An Exploration of Collective Creative Processes on a Collaborative Governance Board, *Group & Organization Management*, 2015, vol. 40, no. 4, pp. 500–528.
[15] T. Brown, *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. HarperBusiness, New York, 2009.
[16] E. Chong and X. Ma, The Influence of Individual Factors, Supervision and Work Environment on Creative Self-Efficacy” *Creativity and Innovation Management*, 2010, vol. 19, no. 3, pp. 233–247.
[17] A. Nakajima, M. Bawiec, K. Nakayama, and H. Horii, The Development of APISNOTE a Digital Sticky Note System for Information Structuring: Case study into the innovation workshops at i. school, The University of Tokyo. *14th International Conference for Asia Digital Art and Design*, 1–2 November 2016, Bali, Indonesia. 2018.
[18] T. M. Amabile, Social psychology of creativity: A consensual assessment technique., *Journal of personality and social psychology*, 1982, vol. 43, no. 5, p. 997.
[19] C. Freeman, J. Son, and L. B. McRoberts, Comparison of novice and expert evaluations of apparel design illustrations using the consensual assessment technique, *International Journal of Fashion Design, Technology and Education*, 2015, vol. 8, no. 2, pp. 122–130.
[20] J. A. Plucker, J. C. Kaufman, J. S. Temple, and M. Qian, Do experts and novices evaluate movies the same way? , 2009, *Psychology & Marketing*, vol. 26, no. 5, pp. 470–478.
[21] M. A. Hearst, Text Tiling: Segmenting Text into Multi-paragraph Subtopic Passages, *Computational Linguistics*, 1997, vol. 23, no. 1, pp. 33–64.
[22] R.W. Zmud, R.J. Mejias, B.A. Reinig and I. M. Martinez-Martinez, *Participation Equality: Measurement Within Collaborative Electronic Environments-A Three Country Study*. Purdue ciber working papers. 2001.
[23] J. R. Mesmer-Magnus and L. A. DeChurch, Information sharing and team performance: A meta-analysis., *Journal of Applied Psychology*, 2009, vol. 94, no. 2, pp. 535–546.

[24] T. B. Ward, Creative cognition, conceptual combination, and the creative writing of Stephen R Donaldson, *American Psychologist*, 2001, vol. 56, no. 4, pp. 350–354.
[25] J. M. DiMicco, K. J. Hollenbach, A. Pandolfo, and W. Bender, The impact of increased awareness while face-to-face, *Human–Computer Interaction*, 2007, vol. 22, no. 1–2, pp. 47–96.
[26] T. Kim, A. Chang, L. Holland, and A. S. Pentland, Meeting mediator: enhancing group collaborationusing sociometric feedback, *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, San Diego, CA, USA, 2008, p. 457-466.
[27] E. J. Wisniewski and D. Gentner, On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning, *Advances in psychology*, 1991, vol. 77, pp. 241–284.