

# A Quantitative Analysis of Collective Creativity in Playing 20-Questions Games

Wen Dong, Taemie Kim, Alex “Sandy” Pentland  
20 Ames Street, E15-383  
Cambridge, MA, 02142  
(wdong, taemie, sandy)@media.mit.edu

## ABSTRACT

Creativity is an important ingredient in problem solving, and problem solving in turn is an important activity for both individuals and societies. This paper discusses our novel approach of discovering the structure of problem-specific creativity with statistical methods, and mapping the interaction patterns of group processes to their performances through the discovered creativity structure. Our discussions are based on a lab study data set using the meeting mediator system through which we collected objective quantitative data. We hope our findings and quantitative approach could be applied to many other real-world problem-solving processes and to helping people.

## Author Keywords

computer supported/mediated creativity, empirical study

## ACM Classification Keywords

H.5.2 User Interfaces: Theory and methods

## INTRODUCTION

Creativity is exercised by every individuals when they make different attempts and find ways to solve their (hard and new) problems. As a result, people created solutions that had not been known to them. Many researchers believe that the everyday creativity has the same mechanism as the one exercised by the great geniuses in the history.

In this paper, we will discuss the creative process, and the relationship between creativity and performance of solving a 20-questions game by a group of people. We will base our discussion on the **meeting mediator data set** [9], so that we can reason the group problem-solving strategies and quantitatively analyze the relationship among the group performances, the group creativity and the group dynamics. We note that the group brainstorming and decision making processes in the meeting mediator data set are representative of the group processes in many task groups.

The **meeting mediator** is a system that detects and displays

group dynamics. Its goal is to quantitatively measure the group’s interaction pattern and provide real-time feedback to promote change in the group’s behavior. The meeting mediator system consists of two components: the **sociometric badge** and a mobile phone. The sociometric badge is a business-card-sized embedded device, that understands the behavior and interaction of the sociometric badge user by collecting the audio and motion information of him and interact with other sociometric badges through radio, IR and Bluetooth channels. It is the input and the brain of a meeting mediator. The phone displays group interactivity level and participation balance which is transferred from the badges via Bluetooth connection. A more detailed description and pictures of a meeting mediator can be found in the work of Kim et al. [9].

The meeting mediator data set involves over 40 group problem-solving processes in different configurations. Our goal in collecting this data set is to understand the relationship between the task group dynamics and the task group performances, so that we can design automated tools to estimate and improve group performances using signals that can be reliably detected and estimated with embedded-device hardware/firmware.

The two tasks in the group processes are two 20-questions games [1]. We split each task into two parts: Each group of four persons is first required to brainstorm as many ideas as possible that is compatible with a partially finished game — a list of 10 yes/no questions and the corresponding answers — in eight minutes. The group is then required to ask as few questions as possible through its leader to get the answer. In order to solve a 20-questions game by asking the fewest number of questions, the group has to construct each question carefully based on its estimation of the items that satisfy the answers to all previously asked questions. As a result, using the two-parts form for the tasks both enforced a good structure in the group problem-solving processes and gives us extra semantic clues on why some groups asked fewer questions than others to accomplish the tasks.

The data about each group process in the meeting mediator data set consist of: (1) the features about the group dynamics and captured by the **sociometric badges** [11] worn by the experiment subjects, in particular at each sample time whether an individual subject is speaking, how he moves his body, and how he orients his body in relation to the others; (2) the facts related to the group performances, including the ideas generated by each team/individual in the brainstorm-

ing sessions, the times spent for every single groups to generate every single questions; and (3) the surveys, in particular those about the subjects' own opinions on the group dynamics and group performances. We neither audio-recorded nor video-recorded any of the group processes, following the COUHES guidelines [2] to protect the privacy of the experiment subjects.

Each group was instructed that collaboration was important to group performances prior to its tasks. The task-groups were randomly assigned one of four labels: A, B, C and D. To test the effect of feedback on the groups' behavior, the members in the groups labeled with C or D were not allowed to see the feedback of the meeting mediators. In the following, we will simply refer the groups labeled with C or D as groups without meeting mediators.

## 20-Questions Game

The current paper will argue that the 20-questions game is amiable for group problem-solving, that the performance of a task-group in solving this game could be reliably estimated from the group turn-taking dynamics, and that a meeting mediator could direct its user to a high level of participation based on its understanding of the group process. Before we inspect the groups in the meeting mediator data set and reason how they could improve their performances, let us make a plan on how we would like to solve a 20-questions game with the best possible performance by ourselves.

In a 20-questions game, the answerer has a thing in his mind, and the questioner is responsible for pinpointing the thing with as few yes/no questions as possible. By information theory, a good question should eliminate half of the remaining candidates that are compatible with the previous question/answer pairs. Since the questioner does not know the set of all compatible candidates a priori, he has to choose a good question based on his sampling of the set. Thus the more similarly the answerer and the questioner sample the candidate sets, the better performance the questioner will have.

We can also put the previous paragraph in the language of probability: The answerer samples a thing with a probability measure  $\mathbb{P}$ ; The questioner reduces half of probabilities at each step assuming he can, or  $\mathbb{Q}(S_{t+1}) = \frac{1}{2}\mathbb{Q}(S_t)$  where  $S_t$  represents the set of all compatible candidates at step  $t$  and  $\mathbb{Q}$  is the probability measure used by the questioner; The closer  $\mathbb{P}$  is to  $\mathbb{Q}$ , the better performance the questioner will have; The questioner samples the set of compatible candidates according to a diffusion process implicit in his mind

$$u_t^{(i+1)} \sim \mathbb{Q}(\bullet | \{u_{t' < t}^{(j)}\} \cup \{u_t^{(1)}, \dots, u_t^{(i)}\}) \uparrow S_t$$

and ask questions based on his samples

$$S_{t+1} = S_t \cap \chi(\{u_{t' < t}^{(j)}\} \cup \{u_t^{(1)}, \dots, u_t^{(m_t)}\})$$

where  $u_t^{(i)} \in S_t$  is an item that the questioner sampled at time  $t$ , and the characteristic function  $\chi$  is what the questioner subsequently asked. Solving a 20-questions game with a group of people provides wider diffusion, larger sample

sizes at each steps, and better questions. When the questioner do not know the  $\mathbb{P}$  in the mind of the answerer, his safest bet is the uniform distribution over the set of all things, and he is expected to pinpoint the thing in  $\log |S|$  steps.

## Literature Review

While a long research history and a large body of literature are associated with understanding group-process dynamics and improving group performance, this research area is often suffered from the high cost of collecting comprehensive and publicly available data sets. The "old-fashioned" research results in social sciences have grains of truths, but they are often imprecise and characterized by the lack of analytical nature. For example, while we all know that a good participation could serve to mix the ideas of the individuals and to generate better group ideas for certain problems, we do not totally understand why and in which types of problems good participation leads to good performance, how to identify good participation, and how the characteristics of good participation can be reliably computed from automated tools.

Some "traditional" work on group brainstorming performance and group decision-making performance is given in the following. Hall and Watson [7] demonstrated that the performance of a group is noticeably affected by the understandings from its members on what is a productive group process, and that the group performance could be noticeably improved by just instructing the group members to be more participative. According to Hall and Watson, a more productive group and a less productive group benefit comparably in problem-solving from choosing the best answers among its members; However, a more productive group is more likely to generate group answers that are better than the individual answers by reconciling the differences among its members with win-win strategies and through 'aha' experiences. Wilson et al. [15] observed several tens of group processes in solving two versions of the 20-questions game from social scientists' point of view. They noted that (1) groups solve significantly larger proportions of the games than individuals, (2) the questions asked by groups work increasingly better than those asked by individuals as a game proceeds (and thus becomes harder), (3) a pair of strangers generate more (unique) ideas — that are compatible with a given list of yes/no questions and their answers — than a pair of friends, and a pair of friends generate more ideas than two individuals working alone. Many issues related to the lacking of participation, such as social loafing and production-blocking, have been discussed by various researchers [12, 8, 10].

Some work on the small-group dynamics that is amiable to analytical approach is listed in the following. Bales [4] noted that a general task-group process is interwoven with speaker turns related to the task functions and those related to the socio-emotional functions, and a satisfactory task-group process strikes a balance between the two types of functions. Zancanaro et al. [16, 6] showed using the *Mission Survival Corpus I* that the different task roles and socio-emotional roles originally proposed by Bales can be reliably detected by statistical algorithms, and the statistics about the roles are

related to the group performances. Conversation and discourse analysis provide helpful observations and examples [3, 13], so that the features and structures of conversational group processes can be figured out by experiments and simulations.

### Plan for the Paper

The rest of the paper is organized in the following way. Section gives the mechanism of brainstorming and the formula to predict decision making performance from the corresponding brainstorming performance, both referring the meeting mediator data set. Section shows how the group dynamics collected by the meeting mediators determine the brainstorming performance and consequently the decision making performance. The effects of the meeting mediators on the group processes in the data set are discussed in Section . We will conclude this paper and mention our ongoing work in Section .

### BRAINSTORMING & DECISION MAKING

In this section, we will discuss how the groups in the meeting mediator data set came to new ideas in the brainstorming sessions, and how the groups subsequently made decisions based on their brainstorming results. We will defer our discussion of the group turn-taking behavior until the next section.

Recall that the task of the groups is to solve two 20-questions game variants. A group will go through a brainstorming session and a decision making session in each game. The group is first given a partially finished game in terms of a list of 10 yes/no questions and their answers, and is required to brainstorm as many items as possible that are compatible with the given list of question/answer pairs in an 8-minute brainstorming session. The group is subsequently required to complete the partially finished game by asking as few questions as possible in a decision-making session. By splitting a 20-questions game into two sessions, we impose a problem-solving structure on group processes.

### Brainstorming

A brainstorming session such as one in the meeting mediator data set can be viewed as a diffusion process. The subjects in the session randomly hit new ideas based on the ideas they have recollected due to familiarity or accessibility and the ideas that they have already considered. The ideas are then checked against the given constraints — a partially finished game represented by a list of yes/no questions together the corresponding answers. If the group hits one idea, it will be more likely to hit other closely related ideas. From this perspective, the set of all ideas that are compatible with the given constraints are related with each other according to the accessibility from one idea directly to another idea, and the inspected ideas of a group diffuse on the network of all ideas. Based on the diffusion argument, a group of people working together will statistically generate more ideas than the same people working individually assuming all other conditions are the same, since in the former configuration people can inspire each other and thus the edges connecting different items are more accessible. Based on the same argument, a

group of interacting strangers will statistically hit more ideas than a group of interacting friends, since in the former configuration the different backgrounds of the strangers imply more “unexpected” edges from items to items [15].

With the knowledge of which group hit which ideas, the structure of the space of brainstorming ideas can be computed: As we have discussed previously, a group either generated all/most elements in a set of closely related items or none/few in this set. In other words, there is an equivalence relation on the set of all items for this group. Thus the set of items generated by a group is the union of some sets of statistically related items (which forms equivalent classes) from the perspective of the group, so is the set of items not generated by the group. When we change our perspective and inspect how the items were generated by the groups, rather than how the groups generated the items, a distance among the items can be defined: Two ideas share an obvious relationship and are thus close to each other if most of the groups generated either both or none of them; Two ideas are related by uncommon relationships and are thus far from each other if many groups generated only one of them. The space of ideas, represented by the set of ideas and the set of relationships among the ideas, can subsequently be mapped into a familiar structure, such as a Euclidean space, a dendrogram or a graph, and be inspected.

We will use task 2 in the meeting mediator data set to illustrate the above method, as well as the structure of the space of brainstorming ideas. The partially finished game of the second task of the meeting mediator data set, represented by a list of yes/no questions together with the answers, is given in Table 1, and the dendrogram of all ideas hit by the groups is given in Fig. 1.

In the meeting mediator data set, 52 groups have their task 2 brainstorming outcomes recorded, and they generated 92 distinct items in total. The overall result can be represented as a  $92 \times 52$  matrix, with the matrix rows corresponding to distinct items and the matrix columns corresponding to different teams. The entry that lies in the  $i$ -th row and in the  $j$ -th column is either 0 or 1, representing group  $j$  did or did not generate idea  $i$  in its brainstorming.

The two characteristics of the dendrogram in Fig. 1 are listed below. They have implications on how a 20-questions game is played, and how the game-play can be modeled, as will be discussed later. First, the brainstorming ideas of task 2 in the meeting mediator data set form small clusters. Of these clusters, some consist of ideas that truly go together (“stapler” and “hole puncher”); and a few — “clusters invented by a clever group” — consist of ideas that were hit by a single group. Second, the small clusters are well separated from each other. As a result, the idea set of a partially finished game can be split into a yes-set and a no-set equally well in many ways, assuming the corresponding yes/no questions can be found. The well-separatedness of the small clusters also means that which cluster of ideas a group is going to hit next is more or less unpredictable.

**Table 1. Task #2 Constraints: Thing**

1. Is it used for entertainment?.....	No
2. Would you give it as a gift?.....	No
3. Is it a tool?.....	Yes
4. Is it smaller than a shoebox?.....	Yes
5. Would you find it in a toolbox?.....	No
6. Does it use electricity?.....	No
7. Is it assembled?.....	Yes
8. Is it meant to get wet?.....	No
9. Does it have metal parts?.....	Yes
10. Is it used indoors?.....	Yes

If we define an Euclidean (i.e., a 2-norm) distance between the brainstorming ideas and subsequently do eigenvalue decomposition on the resulting distance matrix, we will find that the largest eigenvalue accounts for more than 80% of variance, the corresponding eigenvector sorts the ideas based on the frequencies they were hit, and the rest eigenvalues are relatively equivalent to each other. Thus the observation from principle component analysis coincides with the second characteristic we just described based on the dendrogram.

Known the structure of the brainstorming ideas, we could be able to inspect how a group hit new ideas as a function of time, as well as how a group of interacting strangers take unexpected paths and statistically hit ideas faster than a group of interacting friends or a group of people working alone.

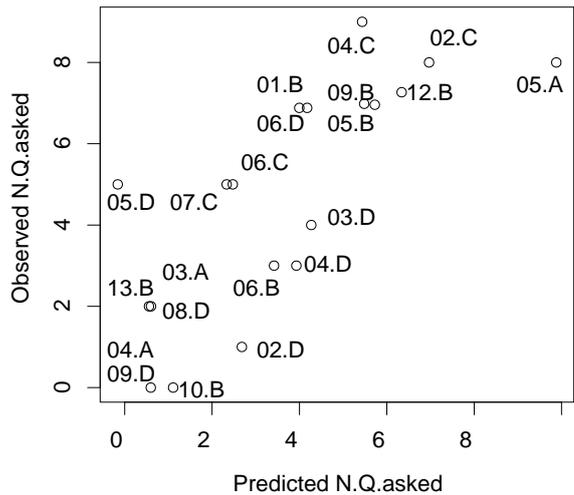
**Decision Making**

The brainstorming session of a task gives a group a considerable amount of time (8 minutes out of 8+10 minutes) in sampling the set of candidate items. It also implicitly prepares the group on the subsequent questions to ask in the next session in order to finish the task. The following lists several strategies that a group can take in exploiting its brainstorming outcome.

A simple-minded approach is to assume that the brainstorming outcome contains the answer, and to conduct a “binary search” on all ideas generated in the brainstorming session. The shortcoming of this approach becomes apparent when the answer turns out to be outside of the brainstorming outcome: The information contained in the additional question/answer pairs is vague, and it needs a considerable amount of efforts to be processed — If the thing is neither a donut nor a bean, it could be not square, not edible, hard, larger than a golf ball, and so on.

A better strategy is to treat the brainstorming outcome as a skeleton of the space of all items compatible with the partially finished game, and to treat the compatible items outside of the brainstorming outcome as the flesh that grows on the skeleton. Thus the question of the group is not “is the thing either a donut or a bean?”, but rather “is the thing edible?” randomly chosen among many choices. The improved strategy has the same performance as the simple-minded one when the answer is in the brainstorming outcome and it im-

Task 2 predicted and observed decision-making performances  
(only for groups that the answer was not in the brainstorming ideas set)



**Figure 2.** The number of questions asked by a group to solve a 20-questions game increases linearly with the distance from the idea locus of a group to the answer. This number decreases with the reciprocal of the diffusion speed that the group discovers new ideas. The labels in the plot are those of groups .

plies better performance when the answer is out of the brainstorming outcome.

A group can be more advanced-minded to estimate the probability measures that the group and the answerer respectively uses. Thus the group can discuss intensively on the consequences of asking “is it edible?” or “is it larger than a golf ball?” among many others before it finalizes the question.

We will discuss the decision-making performance of a group by two cases: when the answer is in the set of brainstorming ideas of the group, and when the answer is not in.

Based on our analysis, when the answer is in the set of brainstorming ideas, a group can identify the answer with around  $\log_2 |S|$  questions (following the partially completed game), where  $S$  is the set of brainstorming ideas. In the meeting mediator data set, 64% and 57% of the groups identified the correct answers with  $1 + \log_2 |S|$  questions in task 1 and task 2 respectively, and 70% and 84% of the groups identified the correct answers in  $2 + \log_2 |S|$  steps.

When the answer is not in the brainstorming ideas, the group will use around  $\log_2 |S|$  steps to realize this fact, and then it will use additional  $d(S, n)/|S|^{0.5}$  number of steps to reach the answer (cf., Fig. 2).

**INDIVIDUAL PARTICIPATION LEVELS & GROUP PERFORMANCES**

We have discussed how a 20-questions game is played by either an individual or a group, and how the performance at playing a game can be estimated from the result and performance of the preceding brainstorming session. We will continue to discuss how the group performance at playing such



storming.) This can be explained by our observation that a group uses most of its time in enumerating and testing ideas, and that our human subjects have comparable skills.

The sum of the fractions of time that the group members take the *giver role* or the *protagonist role* during a group brainstorming session is also a good linear indicator of the group brainstorming performance. This means that the brainstorming performance is explained by the long speaking-turns, which are supposed to be the places where the enumeration and validation of ideas really happen.

All the linear regressions in the above two paragraphs were carried out after we had taken out two data points (out of a total of around 40 data points for each regressions): One group had an average number of 2.5 simultaneous speakers during each of the 8 minute brainstorming sessions, and it only came out around 5 ideas and around 3 unique correct ideas in each sessions. Another group had an average number of 0.4 simultaneous speakers, and it came out around 32 ideas and around 16 unique correct ideas in each sessions. As a comparison, a normal group had an average of 1.2 simultaneous speakers, and it came out around 16 ideas and around 8 unique correct ideas in each sessions. Due to the lack of audio/video recordings, we do not know whether the two group processes proceeded in the same way as the other group processes in the data set.

Leaders are reported to have impact on the task group performance. Due to our need to investigate the characteristics and roles of leaders, each group was instructed to choose a leader after a practice task that precedes and is similar to task 1 and task 2. The group was told that only its leader can ask questions in the following decision making sessions on the behalf of the whole group.

The leaders in the meeting mediator data set generate more ideas than the other group members in around 50% brainstorming sessions. They generated around 7.8 ideas on average for each brainstorming sessions, while a normal person generated around 4.5 ideas. and speak more and longer. The leaders spoke, took the information giver role and took the protagonist role more than the other group members in about 55% brainstorming sessions. Their speaking and turn-taking tendencies are consistent in both of their brainstorming sessions. Thus if we believe that the human subjects' performances are consistent in the practice task, task 1 and task 2, we can conclude that a task group pick its leader from those who perform well and participate well.

### IMPROVING GROUP PERFORMANCE WITH THE MEETING MEDIATORS

We have discussed the structure of the idea space of a partially finished 20-questions game in the meeting mediator data set, the implications of this structure to the subsequent decision making sessions, and how the actions and interactions of the group members are related to the brainstorming performance of this group. We will proceed to discuss how the meeting mediators improved the group performances.

### Task 2 discussion intensity vs. outcome

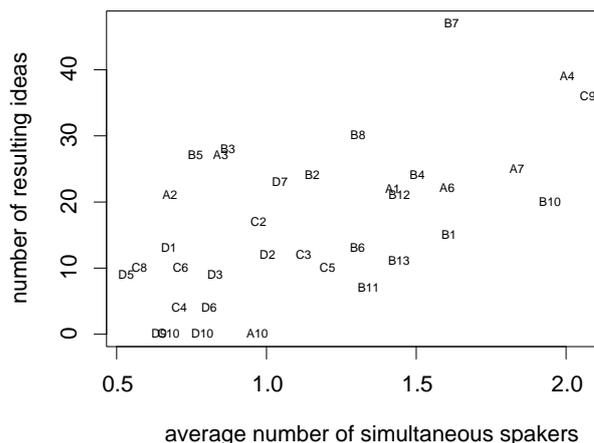


Figure 3. Brainstorming outcome vs. discussion intensity.

By encouraging interaction and engagement, the meeting mediators made noticeable improvements in regularizing the group behavior in the task-group processes without sacrificing performance, and the regularization pays off in performance when the task becomes harder. The more regularized and predictable group behavior has merits in management. In the experiments, the groups equipped with meeting mediators have their performances more linearly-predictable from the interactions in them ( $R^2 \approx .40$  for groups with meeting mediators and  $R^2 \approx .20$  for groups without, cf. Fig. 3). Indeed in performing task 2, even the groups not equipped with meeting mediators intrinsically had more interactions among their members, and their performances are more predictable from the amount of interactions in them. Recall that the overall R-squared statistic increases from .32 to .35 when the groups coped with the harder task (task 2). The meeting mediators also made the group members more considerate and the decisionmaking faster, as was discussed by Kim et al. [9].

A new version of meeting mediator is under development that systematically exploits the group-process mechanism that we discussed in this paper and that could significantly improve the task-group performance.

### CONCLUSIONS

In this paper, we discussed the brainstorming process and the decision making process in the meeting mediator data set, as well as our efforts to facilitate these processes. Specifically, we gave a model on the relationship between the brainstorming performance and the decision making performance, and a model on the relationship between the group process statistics and the group performance. The processes in the data set are characteristic of the learning, planning and searching processes in many cases of the (collective) intelligence that we know of. We have several pieces of future work in our mind. First, we would like to relate the structure and performance of the task-group processes in the meeting media-

tor data set with those in other data sets, attempting to find some general laws of the collective intelligence. Second, we would like to find out the mechanism that intense discussion leads to better group performances. We are currently unable to answer this question due to the restrictions in our data set.

## REFERENCES

1. A description of the twenty questions game. [http://en.wikipedia.org/wiki/20\\_questions](http://en.wikipedia.org/wiki/20_questions).
2. MIT Committee on the Use of Humans as Experimental Subjects. <http://web.mit.edu/committees/couhes/index.shtml>.
3. J. Maxwell Atkinson, editor. *Structures of Social Action (Studies in Conversation Analysis)*. Cambridge: Cambridge University Press, 1985.
4. Robert Freed Bales. *Interaction Process Analysis: a Method for the Study of Small Groups*. Addison-Wesley Press, 1950.
5. Peter J. Burke. Participation and leadership in small groups. *American Sociological Review*, 39:832–843, 1974.
6. Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *ICMI*, pages 271–278, 2007.
7. Jay Hall and W. H. Watson. The effects of a normative intervention on group decision-making performance. *Human Relations*, 23(4):299–317, 1970.
8. S. J. Karau and K. D. Williams. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65:681–706, 1993.
9. Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Pentland. Meeting mediator: Enhancing group collaboration and leadership with sociometric feedback. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2008.
10. Bernard A. Nijstad, Wolfgang Stroebe, and Hein F. M. Lodewijkx. Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, 39:531–548, 2003.
11. Daniel Olguin Olguin, Benjamin N. Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Trans. Syst., Man, Cybern. B*, 39(1), 2009.
12. A. F. Osborn. *Applied Imagination: Principles and Procedures of Creative Problem Solving*. New York, NY: Charles Scribner’s Sons, 1963.
13. Emanuel A. Schegloff. *Sequence Organization in Interaction: A Primer in Conversational Analysis*, volume 1. Cambridge: Cambridge University Press, 2007.
14. Harry L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III*. John Wiley & Sons, 2002.
15. David Sload Wilson, John J. Timmel, and Ralph R. Miller. Cognitive cooperation: When the going gets tough, think as a group. *Human Nature*, 15(3):1–15, 2004.
16. Massimo Zancanaro, Bruno Lepri, and Fabio Pianesi. Automatic detection of group functional roles in face to face interactions. In *ICMI*, pages 28–34, 2006.