

Applying Task Classification to Natural Meetings

Karen Ward

Oregon Graduate Institute of Science & Technology
wardk@cse.ogi.edu

Catherine R. Marshall

Collaborative Technologies
crm@collabtech.com

David G. Novick

Oregon Graduate Institute of Science & Technology
novick@cse.ogi.edu

Technical Report No. CS/E 95-011

Department of Computer Science and Engineering
Oregon Graduate Institute of Science & Technology

P.O. Box 91000

Portland, OR 97291-1000 USA

July 17, 1995

Abstract

This paper considers the problem of understanding the conversational tasks of participants in naturally occurring meetings. Task classification is important for, among many reasons, system design but is an exceptionally difficult problem. In our study, we reviewed some prior approaches to task classification and extended a laboratory-oriented scheme for use in natural conversations. We applied the extended scheme to a variety of meeting conversations, including two entire meetings. The data and our experience indicate that task classification does capture significant aspects of the interaction but that there are problems with this—and likely any—method.

Introduction

Previous studies of technology-based systems designed to support real-time group work show that the appropriateness and effectiveness of such systems vary as a function of the type of task being performed (e.g., Marshall & Novick, 1995; Nunamaker et al., 1991). In order to study the effects of task as a variable in a more systematic manner, we sought a domain-independent method of categorizing tasks for situated group work.

This study grew out of a larger examination of the factors that contribute to the effectiveness of communications and communications technology for collaborative work; our ultimate goal—not addressed in this report—is to provide design requirements for computing and telecommunications systems that will effectively support real-world group work. In an earlier study (Marshall & Novick, 1995), we gathered traditional performance and subjective satisfaction measures in a laboratory experiment comparing the effectiveness of face-to-face, audio-only (telephone) and audio+video (multimedia) communications. Two-person teams were asked to perform two cooperative tasks: a construction task, and a problem solving and negotiation task. For the construction task, teams were more effective in the face-to-face and audio+visual conditions. For the problem solving and negotiation task, teams were more effective in the audio-only condition. As the next step in our research program, we sought to relate the results of our work and that of others to the activities seen in naturally occurring groups working on real tasks in their usual environment.

In particular, we sought a conceptual model that would permit us to relate formal task studies to informal group activities in a way that illuminates task-related differences in communication demands. Such a model would allow us to draw upon the wealth of knowledge about small group processes and small-group communicatory patterns (e.g., McGrath, 1984) to support the design, evaluation, and introduction of new technology to support unfacilitated natural work groups. What tasks should such a system support? How do those tasks interrelate and flow? Are we using appropriate tasks to evaluate our systems? In introducing new technology, do we have an appropriate match between technology and task? Note that we are not trying to observe current behavior with existing CSCW systems. Instead, we wish to establish a principled basis for determining how such systems should be built and what tasks they should support.

One approach to achieving this goal has been to build computational tools that support particular task activities and then observe their use. Electronic meeting systems (EMS) are typically highly structured and are often used in conjunction with formal meeting facilitation. The University of Arizona GroupSystems environment is an example (Valacich et al., 1991). Designed to reduce process losses brought about by the single-thread nature of face-to-face meetings, GroupSystems provides tools and procedures to support the activities of idea generation, synthesis, and prioritization. Structured environments such as these encourage relatively well-defined and homogenous periods of activity with clearly-defined outputs which are amenable to quantitative analysis.

When they can be used, the structured tools and facilitation of the EMS approach may lead to more efficient group processes. Much real-world group work, however, takes place without formal facilitation or support systems. Informal group activities seem chaotic in comparison to the structured environment of an EMS. The group may shift rapidly between several tasks within a

matter of minutes, sometimes without formally completing any of them. We believe that many meetings will continue to take place in this relatively unstructured fashion, and so we are interested in understanding both the communicative processes that occur in informal group work and the tools that are most appropriate to support that work.

The Extended Task Circumplex

A number of task-classification schemes were available that had been developed for use in laboratory or other structured settings. These seemed to us to miss a number of dimensions of interaction, however. Natural group work typically involves shifting among various low-level tasks to accomplish complex group goals. Thus it is generally impossible to categorize a group session as representing one particular type of task in the same sense that one can categorize a laboratory task. Nevertheless, our larger hypothesis—not tested in this study—is that the task-based differences seen in the relatively pure laboratory tasks do occur in the more free-wheeling situated groups. Although the tasks may shift frequently, the communicative characteristics within a slice of the interaction representing a particular type of task should be similar to those seen in the laboratory tasks. Thus we looked for a means to categorize segments of group work in terms that could be related to laboratory-type tasks.

Rather than create our own task taxonomy, we wished to build on previous research. We wanted a task classification that would enable us:

- To classify situated group activities in a domain-independent fashion;
- To code group activities at a sufficiently large granularity to permit coding at 1-2 times real time;
- To relate our observations both to laboratory studies of group interaction and to other studies of situated groups; and
- To classify group activity in a way that exposes and illuminates task-related differences in communicatory demands.

Furthermore, the list of possible tasks should be small enough to permit reliable coding; when coders have many alternatives available, there may be decreased consistency due to coder lapses as they “forget” about some of the alternatives available to them. (c.f., Pruitt and Lewis, 1975).

We are interested in coding group task at an intermediate level. While most group activities exhibit a hierarchy of goals and tasks, we are interested in capturing the group's activity in episodes small enough to allow us to examine conversational phenomena related to the immediate task. It is our hypothesis that differences in conversational interaction will be seen at the level of the current goal. For similar reasons we will not concern ourselves with “hidden agendas” or private goals.

Thus, we did not want a coding scheme as fine-grained as speech-act coding, nor did we want to consider task at the level of the project or meeting as a whole (e.g., Katz & Tushman, 1989). We needed something that would allow us to consider the meeting as a series of short, coherent episodes and allow us to examine the communicatory activities during each episode.

We considered several coding schemes that have been used by other group process researchers. None seemed to meet our needs. Several researchers have used Bales' IPA (e.g., Hiltz et al., 1990) or Conference Process Analysis (e.g., [12]), sometimes augmented with task-specific categories (e.g., Kanki & Foushee, 1989). These are designed to categorize group process, not task. While it would be interesting to compare and relate process to task, these coding schemes do not capture the phenomena we wish to study.

Other researchers have developed their own coding schemes (e.g., Pruitt & Lewis, 1975), generally for the purposes of characterizing interaction in seen in specific laboratory tasks. The coding method developed by the Olsons and their colleagues (Olson et al., 1991) came closest to meeting our requirements in that it is applicable to situated group work, is at an appropriate level of granularity, and is broadly applicable to the type of meeting they examined (design meetings). Many of their categories are related to the specifics of project-oriented design, however, and we were concerned that they would not fit well with the other types of tasks we wished to consider. In particular, the Olsons' method was developed in the context of software design meetings, using both natural group meetings and artificial groups. Although the Olsons have demonstrated that their method applies to the larger domain of general design meetings, we wanted a domain-independent method that would work with a wide range of meeting types

The model that seemed to fit our needs best was the task type model proposed by McGrath (1984). In this model, tasks are characterized in terms of four general processes—Generate (alternatives), Choose (alternatives), Negotiate, and Execute —further subdivided into eight task types along two dimensions — Conflict/Cooperation and Conceptual/Behavioral (Figure 1). These categories are described in Figure 2. The McGrath classification has been used to characterize structured group tasks in the context of EMS (e.g., Tan et al., 1991; DeSanctis & Gallupe, 1987.)

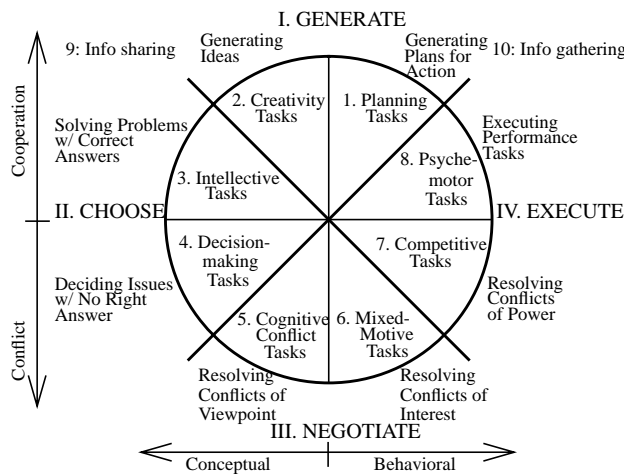


Figure 1. Extended McGrath Circumplex (modified from McGrath, 1984).

Task Type		McGrath Definitions	Extended Definition
I. Generate	1. Planning Tasks	Generating action-oriented plans.	Decomposing an agreed-upon goal into subgoals, often a sequential or hierarchical list.
	2. Creativity Tasks	Generating ideas	Generating alternatives for later evaluation.
II. Choose	3. Intellective Tasks	Solving problems with a correct answer	Performing an algorithmic procedure to produce an answer.
	4. Decision-Making Tasks	Dealing with tasks for which the preferred or agreed-upon answer is the correct one	Evaluating matters of preference.
III. Negotiate	5. Cognitive Conflict Tasks	Resolving conflicts of viewpoint or policy; resolving systematically different preference structures	Expressing or attempting to resolve conflicting policies or values.
	6. Mixed-Motive Tasks	Resolving conflicts of motive or interest; resolving payoff conflicts	Attempting to resolve conflicts of interest, where it appears that one's interests can be met only at the expense of another's.
IV. Execute	7. Contests/Battles	Resolving conflicts of power; competing for victory	Attempt to win a conflict of power or influence.
	8. Performances	Psychomotor tasks performed against objective or absolute standards of excellence or sufficiency	Attempting to perform a psychomotor task to some standard.
Cooperative/Conceptual	9. Information Sharing Tasks	n/a	Reporting or sharing information.
Cooperative/Behavioral	10. Information Gathering Tasks	n/a	Locating information, e.g. through consulting reference materials.

Figure 2. Task Types and Key Concepts of the Task Circumplex

Extending the McGrath Task Circumplex

Preliminary observation of 12 meetings indicated two areas in which the McGrath Circumplex did not seem to capture our observations directly. First, McGrath's Circumplex was intended to be a classification scheme for tasks as designed by the researcher, not a coding scheme for activities as performed by subjects. Second, the meetings we observed included activities that did not fit well into any of the McGrath task types. In this section we discuss these observations and the extensions that they prompted.

Task Definitions

McGrath's Task Circumplex was designed as a classification framework for the laboratory tasks commonly appearing in the group task performance literature. As such, the original definitions embody several implicit assumptions:

- That the defining task characteristics are set by, and thus known to, the experimenter;
- That the task is started, carried out, and completed without interruption; and
- That the task is an indivisible unit.

In laboratory studies or well-facilitated meetings, these assumptions are quite reasonable. The exact task type and goal is generally unambiguous — at least from the researcher's standpoint — and the task has a clearly defined beginning and end. It makes sense to talk about an entire meeting consisting of a single task, especially in the case of laboratory tasks.

In using the McGrath task types as a coding scheme for analyzing meeting activities, we turn these assumptions around. Instead of defining a task and observing subjects' interactions, we categorize portions of a group's interaction in terms of the behavior associated with the task. We describe what people actually do on a minute-by-minute basis, regardless of the intended characteristics of the larger task.

In the natural meetings that we videotaped, a complex high-level task is typically decomposed into subtasks of varying types, often with no explicit declaration of what the participants consider the subtasks to be. Subtasks might be started, interrupted, revisited, carried over to another meeting, or abandoned completely. For example, a meeting whose overall purpose was to decide upon a product marketing strategy (Decision-Making) may contain segments in which the group informally generates a list of alternatives (Creativity Task), debates the merits of the alternatives (Decision-Making), estimates the cost of some of the alternatives (Intellective), adds to the list (Creativity), debates the merits of the new alternatives (Decision-Making) and finally adjourns without making a decision. Should this meeting be considered to have consisted of a single task — Decision-Making — or of a series of tasks? Does it matter that the high-level task was not completed?

To clarify issues such as these, we expanded and reworded the original definitions slightly. The modified definitions are included as a separate column in Figure 2. These changes are intended to allow for the possibility that the task is interrupted or abandoned before completion and to cast the spirit of the original classification definitions into terms that make more sense in the coding context.

Task Types

Substantial amounts of meeting time were devoted to tasks that did not fit well into any of the McGrath task types. In particular, natural groups spend time reporting on work done outside of the meeting, sharing specialized knowledge or expertise with other group members, consulting other sources (e.g., books, notes, or computers; calling someone on the phone), and other activities directed toward ensuring that group members share a common base of information. Occasionally these activities occurred within the framework of another task. Often, however, they did not seem associated with any single clearly-defined task. This was particularly true of reporting and knowledge-sharing activities, which often occurred first in the meeting.

The original McGrath Circumplex does not include an information-sharing task type, which is not surprising in light of the characteristics of the laboratory tasks. The need for establishing a common ground of information is minimized or effectively removed from most laboratory tasks as part of the effort to define the experimental condition. Also, laboratory tasks tend to be short, often only one session, so there is no between-session news or work to report. Because an attempt is generally made to minimize the effect of prior experience on the experiment, participants generally have little background information to contribute.¹ Furthermore, reporting does not seem to

Time Interval	Group activity during interval			Primary task, other notes
	On-task	Meta-task	Task orthogonal	
0:00-0:30	2			Brainstorming
0:30-1:00	2			↓
1:00-1:30	2			
1:30-2:00	2 → 4			Idea evaluation

Figure 3. Task Coding Work Sheet

have elicited much interest as a laboratory task in its own right. To encompass the range of activities seen in natural meetings, however, we needed to categorize information-based activities.

We therefore added two task types to McGrath's original eight: Information Sharing (task type 9) and Information Gathering (task type 10).

Information Sharing encompasses formal reporting, briefing new or previously absent group members, and extended question/answer sessions. This task was placed on the Circumplex in the Cooperative/Conceptual quadrant.

Information Gathering occurs when group members try to locate information by consulting other sources. This includes consulting notes or books, accessing information on computer, telephoning others with questions. This task was placed in the Cooperative/Behavioral quadrant.

Tasks in Situated Group Interaction

We now turn to a description of our experiences in using McGrath's task categories to classify the activities seen in situated group interaction.

We first tested the use of the Circumplex on the protocols collected in our laboratory task study. Our goal at this stage was to refine our coding method and to verify that coders could agree on the categories. Initial tests were encouraging; as expected, the group interactions were easily classified in terms of McGrath task categories. Next we conducted a more formal study in which we analyzed videotapes of several different natural group meetings: employees of a small retail firm reporting on changes required in their departments; MBA students working on a complex marketing simulation problem; academic research groups holding weekly planning and staff meetings; and members of an industrial R&D group holding project team meetings.

In this section we describe the coding method that we developed and report on the results of the formal study.

1. Although in our earlier study (Marshall & Novick, 1995), some participants in a decision-making laboratory task did ground their recommendations in prior experience.

Situated Task Classification

Because our purpose is to classify group activity in a way that will allow us to study communicative phenomena, we needed to find an appropriate granularity for classifying the activities seen in our protocols. The granularity should be large enough to encompass the group's interactive behavior; at the utterance or speech-act level, we are starting to capture individual behavior instead of group behavior. At the same time, it should be small enough to be relatively homogeneous with respect to task; at the meeting level, we are averaging together many diverse behaviors. After some experimentation, we found that 30-second partitions were small enough to be relatively homogeneous while being large enough to allow us to meet our goal of coding at about twice real time. Thus, we partitioned each meeting into 30-second segments and coded the task seen in each segment.

Another approach, which we tested and rejected, is to note the time at which each task shift occurs. We found this to be far more labor-intensive. The coder first must recognize that a task shift has occurred, which may take several seconds as groups do not generally announce task shift explicitly. The coder then typically rewinds the videotape to locate the shift point and note the exact time. When task shifts occur frequently, the coding rate drops dramatically. We realized that we were paying a high cost for an unnecessary level of precision and so decided simply to note that a task shift had occurred at some point within the 30-second interval.

A sample of the coding work sheet is shown in Figure 3. Coders are instructed to enter the task type observed during each 30-second interval in the appropriate activity column (explained below). If a transition from one task type to another occurred during an interval, then both (or all) task types are entered with an arrow indicating the direction of transition. For example, if the group completes brainstorming (a Creativity task) and begins evaluating the brainstormed ideas (a Decision-Making task), the segment is coded: 2→4. For purposes of analysis, we dealt with such transitions by dividing the time interval containing them evenly between the task types coded.

We wanted our coding method to support the eventual testing of hypothesized relationships between the task type and the presence of meta-task and task-orthogonal activities. Meta-task activities are those that address the task process or organization rather than the task itself, as when a group prefaces a meeting by agreeing upon an agenda. Task-orthogonal activities are those that do not relate directly to the task, e.g., one group member offers another a piece of chewing gum and a short discussion about the gum ensues. The work sheet includes separate columns for each of these, plus columns for on-task and sub-goal digression activities.

The coding work sheet also includes an area for noting the primary task and any other observations that the coder feels may be germane. This was used mostly as a diagnostic tool; we found the notes helpful when trying to understand and resolve coder differences.

Data

Our data were drawn from videotapes of several different types of meetings. These included two meetings of three MBA students working on a complex marketing simulation problem ("MBA1" and "MBA2"); the employees of a small retail firm reporting on changes required in their depart-

ments to support a proposed shift in the company business involving one meeting with five participants (“Retail”); two different research groups involving professors, students, and staff holding weekly staff meetings and planning meetings (“Res. A” and “Res. B”); and two meetings of an industrial R&D project team (“R&D1” and “R&D2”). None of the meetings was facilitated, although the employee meeting, the research group meetings, and the R&D meetings each had clear formal leaders. The student meetings had no formal leaders.

Results

We coded two of the meetings (“Retail” and “MBA1”) from beginning to end and took samples from the remaining five. Specifically, we coded one to five 10-minute excerpts from each of the meetings plus two shorter end excerpts (9.5 and 8 minutes) for a total of 18 excerpts covering 355 30-second intervals. Three to five coders (from a total set of 8 different individuals) coded each excerpt.

We determined inter-coder reliability by calculating the mean percentage agreement between all possible pairs of coders for each excerpt. We used an extremely conservative criterion for agreement, in which coders' results had to match exactly. We did not attempt to correct for temporal offsets (discussed in more detail below) or take into account cases in which a coder provided an alternative “close second choice.” As a result, the actual agreement among coders is probably understated. The average reliability score was 0.56 with a range of 0.26 to 1.00. As can be seen from Table I, the scores for one meeting, MBA1, were particularly low. This was one of the two meetings we chose to code in entirety. The choice proved to be unfortunate; the meeting took place in a student lounge and a high level of background noise made the audio track very difficult to understand. This made the meeting difficult to code and, we believe, contributed to the low reliability scores. We excluded this meeting from our subsequent analyses. Without this meeting, the average reliability score for the remaining 14 excerpts increases to 0.63, with a range of 0.41 to 1.00.

Table I. Inter-rater reliability scores

Time	Retail	MBA1	MBA2	Res. A	Res. B	R&D1	R&D2
00:00 - 10:00	0.85	0.34	1.00	0.71	0.53	0.75	0.58
10:00 - 20:00	0.55	0.26	0.83	0.67	0.48		
20:00 - 30:00	0.56	0.33					
30:00 - 40:00	0.41	0.30					
40:00 - 50:00	0.43						
50:00 - 60:00	0.52						

In spite of differences between raters, it was possible to identify a modal response for 93% of all intervals across the 14 excerpts. Based on this modal response, we can begin to look at patterns of task activity in natural group meetings.

We found the meetings to be characterized by different task mixes. For example, participants in the Res. B meeting devoted 17.5 percent of the coded intervals to information sharing, while the R&D2 group spent 68 percent of its time on this task. Likewise, the Retail group spent only 6 percent of their time on planning, while the Res. A and Res. B groups spent 54 percent and 44 percent on planning, respectively. Other task categories show equally strong differences in distribution among the excerpts. These results are summarized in Table II.

Table II. Distribution of Task Types

Task Type	Retail	MBA2	Res. A	Res. B	R&D1	R&D2
1. Planning	6	1	54	44	33	11
2. Creativity	5	0	0	0	0	0
3. Intellectual	0	0	0	0	0	0
4. Decision-making	36	13	0	17	3	4
5. Cognitive conflict	0	0	0	1	0	17
6. Mixed-motive	0	0	0	0	0	0
7. Competitive	0	0	0	0	0	0
8 Psycho-motor	0	0	0	0	0	0
9. Information sharing	45	66	32	17	59	68
10. Information gathering	0	20	0	0	0	0
No task coded ^a	8	0	14	20	5	0

a. Meta-task and task-orthogonal segments.

We can also begin to look at the dynamic temporal qualities of natural group meetings. As Figure 4 reveals, there can be rapid shifting between different types of tasks even within a single 10-minute period.

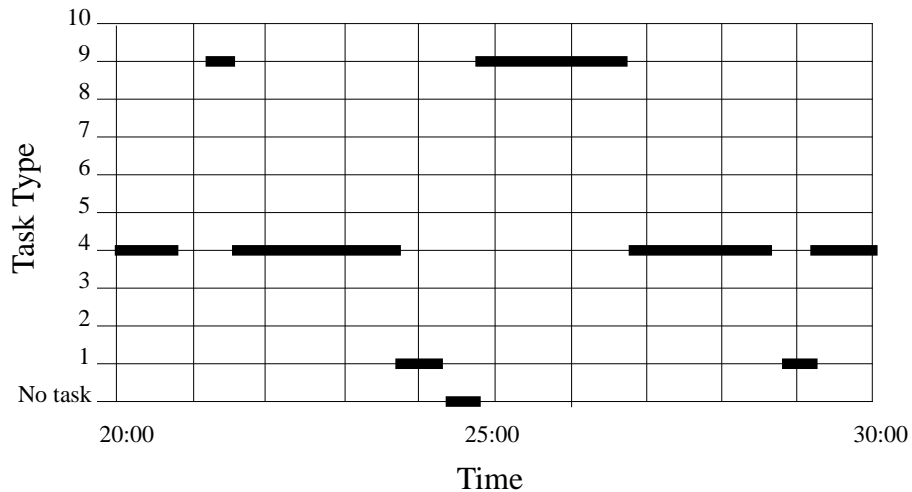


Figure 4. Task distribution for Retail group, 20:00 - 30:00

Discussion

In applying the Extended Task Circumplex to categorize the activities of small natural groups working on situated problems, several issues arose which we believe will be seen in any attempt to apply a standard task categorization scheme to situated group work:

- *Task Granularity.* Even with 30-second time slices, coders sometimes reported as many as three tasks within a single slice. How fine-grained should our analysis become?
- *Group perception of goals and task.* In laboratory studies the task is typically defined for the group; in situated groups, participants may disagree on goals or work at cross purposes.
- *Contextual effects.* Coders' judgements of which task was being performed at a particular time were strongly influenced by their understanding, or lack thereof, of the group's goals and of the domain.
- *Temporal fuzziness.* Coders sometimes detect the same transition but mark the transition point at different times.
- *Categorization.* Coders could not always agree on behavior categorization.

These issues are discussed in more detail below.

Task Granularity

We had difficulty deciding upon an appropriate granularity for our task definitions. Generally, a task will stretch through one or more 30 second segments, and it will be unusual to see more than one or two distinct tasks in one 30-second segment. We found, however, that it was often tempt-

ing to begin treating every utterance as a separate task. This quickly degenerates into a sort of speech-act level coding.

Coding at the utterance level is problematic for two reasons. First, we would like to analyze stretches of conversation for regularities that may relate to observed task-based differences in communicative needs. If the protocol is analyzed at too fine a level, these regularities will disappear. Second, the finer level of granularity is bought at a high price in coder speed. This in turn will sharply limit the amount of data available for analysis.

Group Perception of Goals and Task

Meeting participants may exhibit different points of view about the group goals and current task. In laboratory studies, the task is typically defined for the group. This may obscure the fact that each group participant has an individual view of the problem. For example, McGrath's definition of an Intellectual Task is "solving problems with a correct answer." One of the laboratory tasks we have previously studied, the Subarctic Survival Task¹, in fact has an objectively "correct" or "best" solution. Despite this, some of the subjects solved portions of the problem as if it were a Decision-Making task, one in which the preferred answer is the correct one: after one subject had persuaded the other of one point, they would yield the next point to the other with no debate as if offering a tacit compromise. Although we "knew" that this problem had a correct answer and was therefore an Intellectual task, we decided that these episodes should be categorized as Decision-Making tasks because the meeting participants were treating them as such.

Contextual Effects

Coders' judgements as to which task was being performed at a particular time was strongly influenced by their understanding, or lack of understanding, of the group's goals and of the domain. The group is assumed to be implicitly breaking a single problem into a series of smaller problems contributing to the solution of the whole. In practice, however, groups may explore several alternatives that do not bear a strict hierarchical relationship. They may jump from one task to another, often without verbalizing their reasoning. Lack of domain knowledge may leave observers unsure of the relationship between one activity and the next.

The quality of the video recording also affected the coders' ability to determine task. With poor-quality recordings, the goal of coding at 2 or 3 times real time may not be feasible. Multiple viewings may be needed to establish what was said before coding can begin.

Temporal Fuzziness

In analyzing the data, we realized that coders often detected the same task transition but marked the transition point at slightly different times. We expect that this effect arises for a couple of reasons. First, as there are multiple participants in the conversation, there can be more than one valid interpretation of what is going on at any time in the interaction. Some members of a group may

1. Published by Human Synergetics, Inc.

“move on” to a new task activity while other members still have things to say about the activity that has been going on. In these cases, different coders may pick up on different cues. Second, coders may vary in their sensitivity to the arbitrary boundaries of the coding period. For example, where an activity transition occurs two seconds into a period, one coder may mark this period as a transition between activities and another coder may mark the period as entirely consisting of the second activity. We believe that these effects occur regardless of granularity of the coding period. Moreover, regardless of this fuzziness, coders can be seen to converge on an interpretation, typically within one period of each other.

The technique we have used of selecting a modal response (where available) imposes precision on transitions that are perhaps inherently fuzzy. We are exploring visualization and smoothing techniques that capture the fuzziness of transitions. These techniques may also give us some new insight into the issue of task granularity.

Categorization

We also encountered problems that seemed to arise from the task classification scheme itself. In particular, we found that even when coders agreed on what the meeting participants were doing and when they began doing it, they often could not agree how that behavior should be categorized. The most common disagreements were between adjacent task categories. This is to be expected, since those categories differ on just one of the three underlying dimensions of the McGrath Task Circumplex. However, we also had confusion between task categories 1 (Planning), 4 (Decision Making) and 9 (Information Sharing). It may be that our definition of task categories need further work. Alternatively, it may be that natural group meetings involve a fairly large set of different activities, only some of which fit into our categorization scheme.

Because we observed that coders found it easier to agree on the fact that a task shift occurred and on a domain-dependent description of the group activity, we intend to experiment next with different coding techniques. One approach will be a two-step coding method in which coders will be asked to note only the task transition points and to describe the activity in their own terms. The mapping to a more abstract task representation will be made in a separate pass. We hope in this way to separate the problem of identifying the group activity from the issue of mapping that activity to the Extended Circumplex task types.

Because we speculate that one source of confusion might be the conventional associates triggered by the terms used to describe the task categories (e.g., “planning,” “intellective”), we would also like to experiment with coding using only the three axes that describe the Circumplex: conceptual/behavioral, cooperative/conflictive, and generate/choose/negotiate/execute. We hope that this experiment will help us understand whether the task type names are causing confusion, or whether one of the other distinctions is especially problematic.

Conclusions

The modified task circumplex initially seemed to work well for categorizing the designed tasks used in previous laboratory studies. Applying this same scheme to situated group work was not so straightforward. Even as we were collecting the videotape protocols, we recognized that naturalistic group work typically involves shifting between various “tasks” in order to accomplish complex group goals. Thus we knew that it would be infeasible to categorize a group session as representing one particular type of “task.” This led us to adopt the method of coding time slices in the expectation that we would be able to classify task segments within a session.

We used an extended version of McGrath’s Circumplex of Task Types to categorize the activities of small natural groups working on situated problems. Several issues arose, which we believe will arise in any attempt to apply a standard task categorization scheme to situated group work:

- *Granularity of task.* Despite 30-second time slices, coders sometimes tried to attempt to capture fleeting, utterance-level phenomena.
- *Temporal Fuzziness.* Often coders generally agreed that a task shift had occurred, but varied somewhat in their judgement of where the task transition fell.
- *Lack of agreement on goals and task.* In laboratory studies the task is typically defined for the group; in situated groups, participants may disagree on goals or work at cross purposes.
- *Contextual effects.* Coders’ judgements were strongly influenced by knowledge of the context of prior meetings and of the domain. We do not believe it will be possible to determine task reliably without such background information.
- *Categorization of task.* Even when coders agreed informally as to the group’s task, they often had difficulty agreeing on a mapping to the Extended Circumplex categories.

During the course of this study, our view of task changed dramatically. Initially, we perceived “task” as a grand, overarching, homogenous activity. When we actually observed meetings, we saw instead rapid shifts between low-level activities (cf., Nunamaker et al., 1991). Our observations suggest that “task” entails a complex of goals and activities rather than a single goal or activity. There is usually a nominal task—such as setting sales policies or designing an aspect of a computer interface—that forms a recurrent theme for meeting activities. Episodes of activities directly related to the theme typically are interspersed with other kinds of supporting activities. There may also be implicit tasks—such as group formation or maintenance—that cause other kinds of activities to occur.

Our experience in this study likewise suggests that task context is an important factor in determining observers’ views of the activities. That is, knowing the “nominal” task appears to influence interpretation of the conversational activities. As observers, we impose our expectations of what should be happening on our observations. We are currently designing a follow-on study of other, previously collected protocols, that will test this claim by having naive observers code excerpts of laboratory task meetings.

In spite of these issues, we have been able to use our extended version of McGrath’s Circumplex of Task Types to draw some preliminary conclusions about certain characteristics of natural group work. This is a significant step towards understanding communicative needs and concomi-

tant system design requirements for real-world systems and shows the need for this type of analysis.

In future work we will continue to explore ways to classify group tasks in a fast, robust way that will enable us to understand better task-related differences in communicatory demands.

Acknowledgments

The authors thank Brian Hansen, Brian Korver, David Pennington, Ira Smith, and Stephen Sutton for their contributions to this research. This research was supported by the National Science Foundation, U S WEST, and Intel Corporation.

References

- DeSanctis, G. and Gallupe, R. B. (1987). "A Foundation for the Study of Group Decision Support Systems," *Management Science*, **33**(5), 589-609.
- Hiltz, S. R., Johnson, K., Aronovitch, C. & Turoff, M. (1990). Face-to-Face vs. Computerized Conferences: A Controlled Experiment. Technical Report Research Report Number 12, New Jersey Institute of Technology.
- Kanki, B. G. & Foushee, H. C. (1989). "Communication as Group Process Mediator of Aircrew Performance," *Aviation, Space, and Environmental Medicine*, **60**(5), 402-410.
- Katz, R. and Tushman, M. (1979). "Communication Patterns, Project Performance, and Task Characteristics: An Empirical Evaluation and Integration in an R&D Setting," *Organizational Behavior and Human Performance*, **23**, 139-162.
- Marshall, C. R. & Novick, D. G. (1995). "Conversational Effectiveness and Multimedia Communications," *Information Technology & People*, **8**(1), 54-79.
- McGrath, J. E. (1984). *Groups: Interaction and Performance*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, J. S., & George, J. F. (1991). "Electronic Meeting Systems to Support Group Work," *Communications of the ACM*, **34**(7), 40-61.
- Olson, G. M., Olson, J. S., Carter, M., & Storrøsten, M. (1992). "Small Group Design Meetings: An Analysis of Collaboration," *Human-Computer Interaction*, **7**, 347-374.
- Pruitt, D. G. & Lewis, S. A., (1975). Development of Integrative Solutions in Bilateral Negotiation. *Journal of Personality and Social Psychology*, **31**(4), 621-633.
- Rutter, D. R. & Robinson, B. (1981). "An Experimental Analysis of Teaching By Telephone: Theoretical and Practical Implications for Social Psychology," In G. M. Stephenson and J. H. Davis, eds., *Progress in Applied Social Psychology, Volume 1*, 345-374, John Wiley & Sons, Ltd.
- Tan, B. C.-Y., Wei, K.-K. & Raman, K. S. (1991). "Effects of Support and Task Type on Group Decision Outcome: A Study Using SAMM," In J. F. Nunamaker, Jr., ed., *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, Volume III: DSS and Knowledge-Based Systems and Collaboration Technology, 537-546.
- Valacich, J. S., Dennis, A. R., & Nunamaker Jr., J. F. (1991). "Electronic Meeting Support: The GroupSystems Concept," *International Journal of Man-Machine Studies*, **34**, 261-282.