Patterns of Team Processes and Breakdowns in Information Analysis Tasks

Marcela Borge, Craig H. Ganoe, Shin-I Shih*, & John M. Carroll
College of Information Sciences and Technology & Department of Psychology*
The Pennsylvania State University, University Park, PA, USA
{mborge, szs215}@psu.edu,{cganoe, jcarroll}@ist.psu.edu

ABSTRACT
In this paper we present findings from a laboratory study of teams of three, collaborating to complete a complex information sharing, synthesis, decision-making task. We use interaction analysis, communication analysis, and task analysis methods to identify the primary activities teams engaged in as they solved a complex information dependant decision-making task. These activities serve as the foundation to present findings related to common team problems and patterns of interaction associated with team performance. We found differences between high and low performing teams related to verbal equity and how they shared and synthesized information.

Author Keywords
Information Analysis, Shared Authority, Cognitive Specialization, Team Performance, Information Synthesis, Communication Analysis, Verbal Equity, Information Artifacts

ACM Classification Keywords
H.5.3- Group and Organization Interfaces

General Terms
Human Factors; Experimentation; Performance

INTRODUCTION

Teams often lack the ability to use their interactions as a resource for effective problem solving [2, 3, 13, 14, 16, 17, 27]. In formal learning environments, poor collaborative interactions can lead to lower team learning and task performance outcomes when compared to teams with more effective interactions [2, 3]. The same is true in emergency/crisis work contexts, but with far higher stakes and the potential for catastrophic outcomes [17]. In business contexts differences in processes can be determinants of economic success or an organizations ability to learn [26, 11]. West (2007) argues that successful teams are those that have a greater level of diversity in terms of backgrounds, expertise, and knowledge, and can use those diverse members as resources when engaging in joint team knowledge-building activities. Team knowledge building activity is the act of creating new knowledge that did not exist anywhere in the team prior to collaboration [23]. One person sharing expert knowledge with their team would be an example of information transfer. An example of knowledge building would be the team using that information to draw new conclusions or extend what was previously known.

Our aim is to understand the patterns of activities teams engage in when completing a complex, dynamic task, where the task fluctuates and changes over time, in order to understand how those patterns are associated with performance. Our research questions are as follows: 1) what are the primary activities teams engage in during complex, dynamic collaborative activity (i.e., task analysis and related breakdowns), and 2) what differences exist in the patterns of interaction between high and low performing teams.

Prior research has identified issues of verbal equity and the use of visual representations to mediate understanding as important factors in the learning processes and success of teams [12, 19, 24, 25]. For this reason we chose to comprehensively analyze communication, interaction, development of information artifacts, i.e., visual representations, and compare patterns across teams. By comparing differences between high and low performing teams we aim to identify critical processes associated with performance. Thus we developed a complex scenario to serve as the basis for a lab study where we can control for amount of distributed task knowledge and expertise and meticulously track activity.

We previously developed similar types of scenarios as a means to observe team behavior and inform the design of collaborative tools. We have validated these scenarios with actual fieldwork, and demonstrated that this development process produces tools that improve average performance of collaborative teams [4, 7, 8, 9, 22].

Study design

We used a collaborative information analysis scenario as a means to study information dependent decision-making activity. The scenario needed to fulfill certain requirements

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

in order serve as a representative information dependent decision making task: maximize generalizability, provide diverse but equitable roles, and present a complex information rich, non-linear problem that requires team members to share specialized information in order to make decisions. We decided to develop a complex crime-solving task and utilized a military information analysis scenario as means of informing the structure of an information synthesis task [26].

In order to maximize ecological validity, the scenario is designed to emulate the characteristics and processes of a real police task force team. For example, having investigators that are familiar with the neighborhood and general context of the environment where a crime occurs is crucial to ensuring for an effective investigative team [1, 5]. Consequently the scenario uses local landmarks and crimes that would be more familiar to our participant pool. This is done to provide participants with some inherent expertise, similar to the type of knowledge that real experts might have.

The scenario is divided into three parts and is similar to a proactive investigation tactic employed by police departments when dealing with burglaries, where the ultimate goal to predict and prevent the next crime [5]. Participants must use a mix of inductive and deductive reasoning much in the same way real police officers or information analysts would [1, 20] in order to accurately solve each part of the scenario. Participants must use shared information as a means to accomplish different objectives in each part of the scenario: (part one) narrow down a list of 26 Persons of Interest (POIs) to a list of the eight most likely suspects; (part two) identify thieves for each of four thefts, instigators or accomplices, motives, and whether there were connections among the four thefts; and (part three) predict the thief, time and place of the next crime.

In total, the scenario contains 222 unique pieces of information divided evenly between three information analyst roles: Interview Analyst, Web Analyst, and Records Analyst. Each team member is provided a General Mission Statement and a set of Role Documents. Each set of Role Documents contains unique information that other specialist roles lack. For example, Records Analyst has information on bank transactions, receipts, and class schedules, Web Analyst has information gathered from Facebook, Twitter, Ebay, and other online resources, and Interview Analyst has information gathered from questioning POIs and people they know, and tailing POIs in order to determine regular routines or contacts.

The General Mission Statement is the same for all roles: it contains crime descriptions, explains reliability of the information, and provides "rules" associated with alibis and opportunity to commit the crime. Participants play the same role throughout the scenario, and in order to do well on the task, they need to share their specialized information as a means to draw accurate conclusions.

Pilot tests were conducted to ensure that our participants understood the scenario and had sufficient knowledge of local landmarks and related events. Furthermore, in every trial, we found that participants relied on their knowledge of places and events to make claims. For example, they would make statements like: "Well they could walk from the IST Building to the White Building in 5 minutes; I have done that before."

METHODS

Video data from ten microanalyzed teams, five consistently high performing and five consistently low performing, is the primary data source for the findings presented in this paper. This data is used to unpack the types of cognitive activities and process problems experienced by our teams and whether patterns exist between particular processes and performance. Such statements support that participants used their own knowledge and experience with local landmarks to inform decisions.

Participants

Participants were recruited from a large northeastern United States university. Twenty-two teams took part in the study. Each team was comprised of three participants, for a total of sixty-six participants. The majority of the participants were recruited from undergraduate information sciences and psychology courses. Fifty-three percent of the students were male.

Protocols

Upon arrival, each participant was randomly assigned one of the three roles, which they took on as part of the scenario. In each part of the scenario, the three participants were provided with information particular to their role and were told that only by integrating information and resources would they be able to provide accurate recommendations and successfully complete each task.

The video recorder was set-up to capture the entirety of the lab portion of the study, from the experimenter instructions to behaviors of all team members as they completed each part of the scenario.

Participants had a set amount of time in which to complete each decision-making task: 50 minutes for part one, 45 minutes for part two, and 30 minutes for part three. At the end of each task, participants needed to come to a joint decision and write down a team recommendation or answer for that part of the scenario.

Assessing task performance

Task performance was scored based on accuracy of solutions in each part of the task: total points received/total possible points. There were eight points possible in part 1, 16 in part two, and three points in part three.

Selection of microanalysis teams

Ten teams were selected for microanalysis and their sessions were transcribed and coded. Teams were selected based upon consistent performance across the three parts of the scenario: consistently high performing teams and

consistently low performing teams. Though selection was based on consistency of performance, for the purpose of conciseness these teams are referred to as high and low performing teams in the paper.

In order to identify objective cut offs that could be used as a means of designating performance categories, quartile rank scores were identified for each part of the scenario. Teams were categorized as high, medium, or low teams based on these cut offs. For example, cut offs for part one were 50% accuracy (50th percentile), and 63% accuracy (75th percentile). Teams in the 75th percentile and above were designated with an "H" (high), teams below the 50th percentile received an "L" (low) designation, and teams between the 50th and 75th percentile were designated as mid "M" performers. Teams with the most "H" designations were selected as high performing teams and teams with the most "L" designations as low performing teams. In cases where more than one team had the same frequency of "H" or "L" designations across the three parts of the scenario, average percent accuracy across all three parts was used as a secondary filter.

Transcription of video

Each of the ten selected teams' videos was transcribed following a similar format. Each new speaker utterance and or behavior was numbered, denoting a new "turn". A "turn" ended when a different speaker introduced a new utterance. These turns were then split up into dialogue acts: separate sentences. Compound sentences were split into separate acts as well. The participants were referred to by the role they played and were given pseudo-names in the transcript. Parentheses were used to label nonverbal gestures and events (i.e., leaving the group, making faces, using hands, etc). Brackets were used for codes, time stamps, and notes relevant to the analyses, but not found in the video itself. These transcripts were utilized along with video when analyzing the artifacts and detecting common task errors.

Identifying primary activities

The identification of primary activities was the result extensive qualitative analysis headed by the first and third author and informed by members of our lab. Students were trained to observe and take notes on teams while the experiment was running. The first author developed content logs and notes of behaviors for each of the microanalyzed teams (for description of content logs see [15]). These notes and logs were used to develop conceptual models of the activities the teams engaged in. These conceptual models were then compared to student observations from all of the teams. The third author in collaboration with students then went through the videos again, with transcripts, to identify common team problems. Group discussions between the authors and students were used as a means to verify observations, organize the primary tasks and related activities. This process often included sharing notes, conceptual models, and discussing video segments. This method is similar to methods of interaction analysis [15]. Through this process we identified the primary activities teams engaged in, subactivities, problems encountered by teams in each category, and the relationship between dialogue coding patterns and primary activities.

Coding dialogue patterns

A method of analyzing communication was used that built on previous work [8], but modified to better meet the needs of the current research. Inter-rater reliability for the new coding schema was Kappa= .67, indicating substantial inter-rater reliability [18].

In the coding schema, dialogue is divided into four classes (in bold below) and then broken down into acts. These acts can be compared to other variables such as performance in order to distinguish patterns between successful and unsuccessful teams.

Information Transfer- How new information (existing prior to collaboration) is added
(AI): Add Info- Add new information w/o prompting
(Q): Question- Prompt someone for new information
(R): Reply- Provide new information in response to a prompt
Check Understanding- How previously added info is checked, confirmed, or repaired
(CH): Check- verifying information
(CL): Clarify- clarifying or restating information
(AC): Acknowledge- signaling receipt or understanding of information
Management of Processes- How work is orchestrated
(MN) Management- discussions centered on interactions, how to do the work, and what has been done
(CM) Command for action, order, or instruction that does not take others into account
(RQ) Request for action- posed as a question or indirect prompt (not a question)
Interpretation- How task information is interpreted
(J): Judge- Individual preference, opinion, or claim, with or without deliberation
(RA): Rational that supports a judge (J) or alternative (AT) act.
(AT): Proposing alternative to a (J) OR (RA) act.
(CO): Confirmation- Requesting agreement on a proposed decision
(AG): Agreement- Indicating agreement for prior judgment or decision

Once dialogue was coded, the raw scores for each team were normalized. This was accomplished by calculating the percentage of specific dialogue acts to total dialogue acts. In this way comparisons could be made across teams and we could control for amount of talk. Dialogue act coding was used as a means of looking for patterns of interaction as they occurred in different primary activities within the scenario.

Measuring verbal equity

Verbal equity scores were based on amount of variance of speech turns. Each team's members were designated as high, medium, and low verbal contributors based on

number of speech turns. For example, when calculating scores for overall talk, "percent of total talk" was calculated for individuals by dividing their total number of speech turns by the total number of speech turns for the entire team. The standard deviation of variance was calculated for percentage of total talk across all teams: Mean amount of talk was .33 (33% of talk), SD = .07, (7% of talk). This standard deviation was used as a means to score and categorize teams into three categories of verbal equity.

Verbal equity scores distinguish between patterns of communication: score of 3, or equitable verbal participation, score of 2, or shared authority, and score of 1, or dominance of one team member. The method of calculating scores is detailed below:

h = # of turns of high speaker
m = # of turns of mid speaker
l = # of turns of low speaker
t = Total # of turns
SD= standard deviation of talk

Score of 3: (h/t)-(l/t)<2SD
Score of 2: (h/t)-(l/t)>2SD
(h/t)-(m/t)<1SD
Score of 1: (h/t)-(m/t)>2SD or
(h/t)-(l/t)>2SD and (h/t)-(m/t)>1SD

Each team received three verbal equity scores, one for each part of the task. Average verbal equity scores were calculated for each team and used as a means to compare to outcome measures.

Collection and coding of information artifacts

Multiple sources of data were used to assess and code information artifacts: information provided by participants, information from video and transcripts, and also measures of team accuracy at solving the task.

Shared information artifacts are visual representation utilized by the entire team. They were scored for epistemic fidelity [10] and type of use. Epistemic fidelity (accuracy and completeness) was measured by matching the information needed to solve each phase with the information contained in the artifact.

Type of use refers to the primary function that an information artifact served. After videos were fully transcribed, they were coded to highlight any time shared information artifacts were created or in use. This allowed for the identification of when artifacts led to the creation of other artifacts, when artifacts were referred to or used, and when artifacts were used in conjunction with other artifacts. This information was used as a means to differentiate between note-taking or knowledge-building activities. Note-taking activities describe instances when teams use the information artifact to write down what the team is saying, but no evidence exists of the team pulling information from the artifact to lead to a new conclusion. Knowledge building activities, on the other hand, refer to instances when a team will pull information from an artifact to guide discussion in new directions, ask novel questions, or create new artifacts.

RESULTS

Primary team activities

Six primary activities emerged from the microanalysis: 1) reading/searching intelligence reports, 2) sharing information, 3) synthesizing information, 4) interpreting information, 5) making final decisions, 6) and receiving new information. These activities were not carried out in a linear fashion and patterns of flow from one to the other varied by team and by the strategy they chose as a means of sharing information. For example, teams that chose to proceed through the task by sharing information from one POI to the next were more likely to cycle through the primary activities than those who shared information based on crime location, or with no clear strategy. Thus far, no relationship has been identified between the order of activity flow (i.e., how teams move from one activity to the next) and their performance, but more work remains to be done.

During *reading/searching the intelligence report* participants worked independently as they learned about the task, their role, and their unique role information. The documents our participants received contained information similar to the types of real-world documents that specialists might read prior to engaging in a real-world information synthesis scenario: a brief containing the general problem, situation as it stands, their role in the team, and, in keeping with the scenario, a summary of all of the information their specialized team had gathered to date. During this activity participants familiarized themselves with the intelligence: reading, searching, interpreting, highlighting information, and writing on the documents. Individuals also used information from the intelligence report to create individual information artifacts: notes or visual representations. At this point in the task, team members did not speak to each other.

During this time, the primary cognitive activity was individual knowledge building: developing knowledge individuals lacked prior to the start of the scenario. Individuals were instructed to use the given information to generate preliminary judgments about potential suspects. Individuals read over the information, highlighting information and creating notes of their interpretations. The primary tools used in this category were pens, markers, and separate pieces of paper: pens to underline information, markers to highlight, and paper to keep separate notes. Maps, sticky notes, and calendars were generally not utilized at this point in the activity.

Most individuals spent more time searching and interpreting crime related information (POI information in the role specific documents) than they did searching and highlighting mission goals and objectives, and understanding the problem. After reading through the information and forming an understanding their role,

individuals wrote down their initial interpretations of the information and prepared to introduce themselves and their role to team members.

Once participants finished reading the documents and writing notes they moved on to the next primary activity: *sharing information*. This activity includes all actions where participants discuss preexisting knowledge: knowledge already gathered and represented in the intelligence documents. There were five categories of actions that occurred as part of information sharing activity: 1) sharing information relevant to the crimes, 2) sharing information about the crime-solving task, 3) filtering information, 4) checking/clarifying information, 5) and prompting for more information.

Sharing information relevant to the crimes refers to the sharing of role specific document information, i.e., information about POIs that one role has, but another may lack. This type of information comes directly from the Role Documents and contains no inferences. One example could be, "I have that Tay was in his statistics class on the 5th."

Sharing information about the task refers to instances when participants refer to information contained in the Mission Statement and General Instructions, i.e., information about the crime-solving scenario context. This would include sharing of any of the task rules, goals, or background information contained in the documents participants received. For example, "Well it says here that even if they are listed as attending a class, they can still sneak out for 10-15 minutes depending on what type of class it was."

Information can be shared with a group in one of two ways: though *push* acts and *pull* acts [8]. A push act refers to instances where a team member shares information from intelligence documents without being asked to do so. We code this as an *add information* act (AI). Information can also be shared in response to someone prompting for it; this is referred to as a pull act, or a *reply* act (R) in our coding schema. Prompting for more information refers to instances when participants would ask questions to illicit more information transfer, such as, "Do you have any more information about Dante?"

Participants also checked and clarified transferred information. These types of behaviors include questions and answers that are not aspects of creating new knowledge, but rather clarity and full understanding of previously shared existing knowledge. For example, "Wait, what building did you say Isabel was in", would be an example of a *check* behavior in this category. A clarify behavior that could follow would be, "She was in the White building." Clarify behaviors in a sense are repeated information.

The primary collaborative breakdowns that occurred during *sharing of information* were lack of shared understanding due to imprecise sharing and ignorance of team members' questions and opinions. These breakdowns had negative consequences either with overall team performance or for a particular interpretation or decision act. They were primarily present in teams where one person dominated most of the verbal contributions.

Lack of shared understanding can lead to synthesis problems as false conclusions result from misinterpretations. Even superficial examples where teams do not realize they are talking about different POIs (Cristian vs. Kristin) or different Starbucks café locations could have a large impact on decision-making. This is due to the fact that such information serves as the foundation for synthesis from which a team draws final conclusions.

There were several incidents related to problems with dominance where one team member repeatedly asked for a piece of critical information, but was ignored by a dominant member. Thus key pieces of information were not shared that could have helped the team to connect opportunity or motive to a suspect.

There is no real team knowledge building that occurs within this category. *Individuals* may build new knowledge, but since the information already existed in some form within the team, team members are simply transferring knowledge between members. This shared knowledge is then synthesized to some extent by the team. This is the next main activity that teams engaged in.

During *information synthesis*, teams figured out how to combine the separate bits of shared information to create a collective whole. They figured out how to create a shared understanding of the problem, or a shared representation of the problem. Sub-activities within this category include, 1) discussing strategies for sharing information, 2) summarizing information, 3) creating personal notes, 4) creating information artifacts, and 5) referring to information artifacts.

One aspect of the synthesis category is related to verbalization of strategies for information sharing. These are the rules that teams set up for how they would contribute information, e.g., organizing information sharing according to POIs' first names and going in alphabetical order, turn taking, etc. Many teams came up with an agreed upon order to share information and then synthesized information in a personal or shared artifact. However, relationships between strategy use and effectiveness have yet to be examined in the dataset.

Summarizing information was another sub-activity in this category. Summarizing was more frequent when teams did not create detailed externalized shared representations as information artifacts. Thus teams would provide a summary of all of the information the team had already discussed: this included transferred information and inferences made from transferred information. For example, "Okay so Tay and Isabel where both at work as were Edgar and Frank. And we decided that Tay and Isabel could not have committed the crime, but we're unsure of Edgar and Frank." Summarizing information was also used as a

means of checking the accuracy and completeness of information artifacts or as a means of determining the team's current knowledge state.

Many individual team members kept personal notes of the information that other team members presented. However, personal notes even when containing synthesized information, were associated with collaborative process problems. There were instances where individuals wrote down correct solutions, or critical bits of information in their notes, but still failed to share them with the group. Below is an example of this type of problem as it occurred in one team. One team member had critical information written down in his personal notes, but failed to share it with the team. The team identified this breakdown during the reflection period following part two of the scenario:

Turn	Role	Utterance
1.	Record	Yeah. Nicky hit York's car. Nowhere does it say George needed his car repaired. I don't have anything on that do you have that?
2.	Interview	No, I have nothing.
3.	Record	Do you have anything for George?
4.	Web	Um, let me look through it again. Um, no, but I told you George looks like a shady character. Oh, yeah (looking at notes), right here- Oh s***! This is all my fault. I did say something about George having really aggressive- It says, "I can't believe Upton f*** up and vandalized my car, I saw him." I think this would have been really helpful.
5.	Record	Oh, well. That's good to know. Aw, well.
6.	Web	Sorry guys.

In this example, the team was comparing the solutions they provided for part two with the correct solutions that the experimenter provided them to reflect on before beginning part three. They were having difficulty understanding how George could have been a thief since they recalled no real mention of him or any motives for committing the crime (turns 1, 2, and 3). In turn 4, the Web Analyst remembers saying that George was suspicious at some point, but realizes he failed to provide his team with a rationale as to why. He had made a note about George, but hadn't shared this information with the team. The information stated that one of the victims of the laptop theft had vandalized his car. That information, he claimed, would have been helpful to solving the crime.

Unlike personal notes, shared information artifacts, by definition were utilized by the entire team. Information artifacts are visual representations developed as a means to synthesize shared information. Both high and low performing teams engaged in elaborate discussions centered on how to orchestrate the creation and development of a shared artifact: what type of representations to create, how to highlight, organize, and filter information, and who should write on the artifact. Not all teams engaged in these types of discussions and preliminary data does not indicate that these discussions lead to higher task accuracy.

One of the most important activities teams engaged in when orchestrating the creation of a shared information artifact was deciding how to filter information. Filtering of information was something that many teams struggled with and that led to problems with synthesizing and interpreting information. Teams used variables related to the scenario, i.e., crime locations, POI relationships, events, etc., as a means to select and organize information depicted on the shared artifact

Once teams created information artifacts they engaged in other subtasks of information artifact related behaviors: they followed rules, ignored rules, added information, created new related information artifacts, and created new unrelated artifacts. Following or ignoring rules refers to whether the information artifact had systematicity, i.e. whether they followed uniform rules of correspondence [10]. For example, if eliminated suspects were to be written in green ink, was this rule applied uniformly. Lack of systematicity was a common occurrence in both high and low performing teams. When symbols and rules were not uniformly followed it made searching for and retrieving information from the artifacts more difficult. This was particularly true of overly complex artifacts such as one created by a consistently low performing team that had so many symbols and rules that teammates had problems keeping track of them and as a result added and interpreted information incorrectly.

Though *interpreting information* was often tied with synthesis, it could technically occur after any of the primary activities: reading intelligence reports, sharing information, or, synthesizing information. Interpretation begins once a judgment has been made. A judgment act is a hypothesis, opinion, or claim, made with or without deliberation. Judgments can be made about crime information, the preexisting information the team has shared about suspects, locations, relationships, etc. Judgments can also be made about how to do the task, i.e., what counts as evidence or alibis, what types of representations are best, etc.

Once a judgment is made teams can respond to them in one of four ways: they can ignore them, reject them, accept them, or they can discuss them. Similar categorizing of utterances has been used when evaluating team interactions in educational settings [2]. Ignoring a judgment means that a judgment is simply disregarded and no evidence exists in transcript that any team member paid attention to. Rejecting a judgment means that teammates refuse to accept the judgment without providing any rationale. For example, they could reply to a judgment by saying, "that's ridiculous", or "you're wrong" and moving on. Accepting a judgment means that a teammate agrees with it without adding any comments, (i.e., "Yeah, you're right", "Yeah, [repeats judgment]), denotes judgment on shared artifact, adds related information, or connects shared information.

Discussing a judgment means that the judgment is not accepted outright, but rather is examined or considered. Examples of discussion type utterances include, "Why do you think he is suspicious", "Just because he has class doesn't mean he has an alibi, he could have snuck out", or "yeah, but you don't have any evidence to support that".

Collaborative process problems related to interpretation included, decisions based on partial information, and faulty hypotheses-guided information gathering. Partial information refers to instances when teams made decisions without having all of the facts relevant to the crime. Faulty hypotheses-guided information gathering occurred when information gathering and, thereby, judgments were biased by pre-mature hypotheses. For example, in part one, some groups hypothesized that all the thefts were connected (a false assumption) and were associated with two competing teams. Therefore, they judged that everyone on the victim's team innocent (another false assumption). Similar events occurred in other groups. This pattern of interpretation was associated with over rationalization: high frequency of rationale acts. Not surprisingly consistently low performing teams had higher frequencies of rational dialogue acts than consistently high performing teams. Groups with patterns of faulty hypotheses-guided information gathering also had problems with synthesis activities: information artifacts were not properly filtered, or information from artifacts was not synthesized to develop new artifacts or knowledge.

After teammates respond to judgments they *make final decisions*: the fifth activity teams engaged in. Final decisions were made with and without confirmation acts. Confirmation acts were speech acts where team members attempted to establish certainty about a judgment. Examples of confirmation speech acts include such questions as, "So, we are eliminating Tay, right?", and "So, yes to Isabel?"

Errors related to making final decisions primarily centered on poor time management. Most of the lower performing groups made hasty final decisions because they ran out of time. These decisions also tended to be made by the dominant member, often written down without any discussion at all. Our highest performing team had one member who took responsibility of time management. This person kept the team focused and encouraged them to share all relevant information rather than spend too much time on one person. In contrast low performing teams spent a great deal of time on a few suspects and therefore did not have time to discuss them all.

Finally, as the team progresses through the crime-solving scenario they *receive new information*. This information could be related to previous information or it could stand-alone. For example, part two of the scenario requires the team to combine new information with previous information from part one. New information can also be unrelated to previous information: to solve part three, the team did not need to look at information from parts one or two, though they were not told this.

Differences in patterns of interaction

Once we had identified the different activities inherent of the task we used these activities to compare the interactions of our consistently high performing groups to those of our consistently low performing teams. This was done by looking at patterns of verbal equity and use of specific dialogue acts.

Verbal equity was assessed for a team's overall talk (as determined by turns of speech) as well as for each of the primary activities (as determined by dialogue acts). As detailed in the methods section, an equity score of 1 denotes domination: one person dominates the group. A score of 2 signifies a shared authority pattern, where two people share most of the talk and one person is significantly lower. A score of 3 indicates a pattern where all three members contribute equally to the talking space and no significant differences exist in frequency of talk between any two members for a specific activity.

There was a significant difference between the verbal equity patterns of overall talk for high performing teams and those of low performing teams, $t(8) = -3.01$, $p < .05$. High performing teams had a mean equity score of 2.7, SD = .58. Whereas low performing teams had a mean equity score of 1.6, SD = .55. However, verbal equity fluctuated depending on the type of activity that teams engaged in, as defined by our dialogue acts and the primary activities. For example, one of the highest performing teams (team 13) had overall equity scores of 3, but when examined closely this overall equity pattern was actually the result of shared dominance in different activities and not equity throughout the differing activities. Shared dominance describes a pattern where or one person dominates one type of dialogue class, i.e., synthesizing information, and another person dominated a distinct class, i.e., interpretation.

Verbal equity was assessed for dialogue acts related to *sharing information, information synthesis, and interpreting information*. The average score across all ten

	Team	18L	21L	17L	1L	8L	15H	7H	14H	13H	2H
	Ave. Performance	25%	28%	29%	36%	39%	50%	54%	57%	57%	60%
Verbal Equity Score	Info. Transfer	3	1	1	3	1	3	3	3	3	3
	Synthesis	2	1	1	1	1	1	1	3	1	1
	Interpretation	1	1	1	1	1	3	3	1	1	1

Table 1. Table of verbal equity scores of the ten microanalysis teams for three of the primary activities. High performing teams (H) are on the right and low teams (L) are on the left.

teams for *sharing information* behaviors was 2.4 with a standard deviation of 0.84. The average verbal equity for *information synthesis* and *interpreting information* were lower. The average verbal equity score was 1.3 (SD = 0.64) for synthesis and 1.4 (SD = 0.84). As Table 1 illustrates, both high and low teams were more likely to have one person dominate synthesis and interpretation activities than they were to have one person dominate information-sharing activities. However there were noticeable differences between high and low teams with regards to how information was shared and synthesized, and how authority shifted depending on the primary activity.

Consistently high performing teams had significantly higher proportions of *push* acts to total information transfer (IT) acts than did our low teams, $t(8)=-2.506$, $p < .05$. Consistently high performing teams also had significantly lower proportions of *check* and *clarify* acts to total IT acts as compared to consistently low performing teams, $t(8)= 2.99$, $p < .017$.

When it came to synthesizing information, consistently high and low performing teams used visual representations to support different functions. Consistently high performing teams used them to record and guide externalized knowledge-building behaviors: they built on prior information, correlated data from multiple artifacts, and used it to come up with new forms of externalized information. In contrast, lower performing teams used artifacts primarily for note-taking purposes: to synthesize verbalized information or to write down solutions. In fact the proportion of artifacts used as knowledge-building tools to total artifacts for consistently low teams was 22.2%. Whereas, the proportion of artifacts used as knowledge-building tools to total artifacts for consistently high teams was 88.9%. The information contained in the artifacts of consistently low performing teams was often disjointed and artifacts were rarely used in conjunction with other artifacts. This led to misinterpretation of data based on incomplete information. Consistently low performing teams externalized shared information, but rarely externalized knowledge-building activity.

There were also differences between the amount of variables high and low performing teams used to organize and narrow down represented data. For example, when creating lists, low performing teams recorded all information that fell under one variable, e.g., date of crime. Low performing teams would then list any shared information related to that date, i.e., where all the POIs were at the time. In contrast, high performing teams used multiple variables to reduce the amount of depicted information from the amount of shared information. For example, they would use the same list with the crime date heading, but only list information related to people who were at or near the scene of the crime at the time of the crime. This facilitated data reduction.

Another difference between consistently high and consistently low performing teams had to do with authority shifts between primary activities. The biggest difference between the high and low performing teams was not in terms of the equity scores related to the primary activities, but rather how dominance varied by person between synthesis and interpretation activities. When an individual dominated the synthesis related activities in a high performing team, a different person dominated the interpretation of information, or decision-making activities. For example if a high performing team had a member contribute 50% of the dialogue acts for synthesis, a different member would dominate around 50% of the interpretation acts. In contrast, a low performing team would have the same person dominate both activities. The difference in shifting authorities is depicted in Figure 1. In this figure the first team, with a verbal equity score of 1, has the same member dominate regardless of activity. The second team, a consistently high performing team with a verbal equity score of 2, has one member dominate synthesis and a different member dominate interpretation. This "cognitive specialization" or shared authority was present in the two highest performing teams out the 22 teams we ran; this pattern was not present in any of the low performing teams.

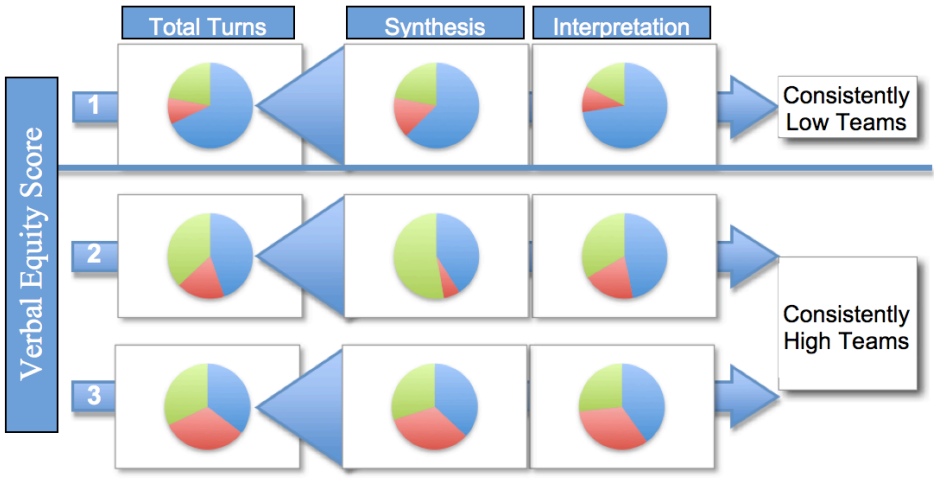

Figure 1. Depiction of most common verbal equity (VE) patterns for the three members of consistently high and low performing teams by type of activity. Low performing teams were more likely to have dominant patterns (VE score of 1). Whereas, high performing teams either displayed shared (VE score of 2) or equitable patterns (VE score of 3).

DISCUSSION

Primary activities

We used qualitative methods to identify the primary activities that participant teams engaged in and problems that arose as they completed the crime-solving scenario. We found that all of our teams engaged in higher order thinking processes as they synthesized and interpreted a great deal of data in a relatively short amount of time. Thus our scenario succeeded in eliciting complex higher-order behaviors.

We identified primary activities as a means to help us develop a better conceptual model of the types of activities we want to support with technology. This approach to design is an effective way to help ensure the effectiveness of technological tools [21].

The identified activities were used as a means to examine differences in patterns of interaction between high and low performing teams. We found differences related to how teams shared and synthesized information, and also found differences in verbal equity patterns.

Sharing information: Push and check/clarify acts

When we looked at how teams shared information, we found differences in the proportion of information added by team members without prompting relative to total information transfer acts. High performing team-members *pushed* at higher frequencies than members of low performing teams. They also had lower frequencies of communication acts related to checking, repeating, and clarifying previously added information. These findings replicate previous findings that equated high *push* acts and low *check/clarify* acts with higher levels of activity awareness [8].

Synthesizing information: Use of multiple variables

There were differences between high and low performing teams with regards to how they synthesized information. Higher performing teams used multiple variables (e.g., time, location, etc.) related to the crime scenario to select information to represent visually on their shared artifacts. Low performing teams synthesized all shared information and did not effectively reduce data as they moved from sharing to synthesizing information. This suggests that identifying variables associated with the specific real-world task is important, as combinations of these variables may be an effective way to filter information and reduce the amount of visualized data that users would have to decipher.

Verbal equity: Shifts in authority

High and low performing teams had very different patterns of verbal equity. Overall, high performing teams were more verbally equitable than low performing teams. However, our study demonstrated that verbal equity can be interpreted in different ways and that the types of activities team members engage in may influence verbal equity patterns. When we evaluated verbal equity based on activity rather than overall speech turns, we found shifts in authority (i.e., who controlled most of the behaviors associated with an activity) for members of high performing teams and none of these shifts for low performing teams.

Findings related to how verbal equity fluctuated by activity may help to explain why verbal equity has been associated with group intelligence [28]. It is possible that teams with higher verbal equity make better use of team members as resources and better integrate whatever specialized knowledge or experience different members bring to the table. However, more work needs to be done in order to determine the relationships between verbal equity and the integration of team member expertise and diverse perspectives.

ACKNOWLEDGMENTS

This project is partially funded by the Office of Naval Research and through the Edward Frymoyer Chaired Professorship.

REFERENCES

1. Alifano, C. M. *Fundamentals of criminal investigation.* New York: Worldwide Law Enforcement Consulting Group, Inc., 2006.
2. Barron, B. When smart groups fail. *Journal of the Learning Sciences*, 12(3) (2003), 307-359.
3. Borge, M., & Carroll, J. Using collaborative activity as a means to explore student performance and understanding. In *Proc the International Conference of The Learning Sciences,* ACM Press (2010).
4. Carroll, J.M., Mentis, H., Convertino, G., Rosson, M.B., Ganoe, C.H., Sinha, H. & Zhao, D. 2007. Prototyping collaborative geospatial emergency planning. In *Proc of ISCRAM (2007)*, pp. 105-113.
5. Chaiken, J. M., Greenwood, P. W., & Petersilia, J. R. *The criminal investigation process: A summary report.* Santa Monica, CA: RAND Corporation, 1976.
6. Chin, G., Kuchar, O.A., and Wolf, K.E. Exploring the analytical processes of intelligence analysts. In *Proc. CHI 2009,* ACM Press (2009), 11-20.
7. Convertino, G., Mentis, H. M., Rosson, M., Slavkovic, A., and Carroll, J. M. (2009). Supporting content and process common ground in computer-supported teamwork. In *Proc. CHI 2002*, ACM press (2002), 2339-2348
8. Convertino, G., Mentis, H., Rosson, M.B., Carroll, J.M, Slavkovic, A. & Ganoe, C.H. Articulating Common Ground in Cooperative Work: Content and Process. In *Proc. CHI 2008,* ACM Press (2008), 1637-1646.
9. Convertino, G., Mentis, H.M., Slavkovic, A., Rosson, M.B. & Carroll, J.M. Supporting knowledge sharing and awareness in distributed emergency management planning. *ACM Transactions on Computer-Human Interaction*, (In press).

10. diSessa, A. Metarepresentation: native competence and targets for instruction. *Cognition and Instruction*, 22, 3 (2004), 293-331.
11. Edmondson, A. C. The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science* (2002), 128-146.
12. Greeno, J. G., & Hall, R. P. Practicing representation: Learning with and about representational forms. *Phi Delta Kappan*, 78 . (1997), 361-367.
13. Hogan, K. Relating students' personal frameworks for science learning to their cognition in collaborative contexts. *Science education*, 83, 1 (1999a), 1-32.
14. Hogan, K. Thinking Aloud Together: A Test of an Intervention To Foster Students' Collaborative Scientific Reasoning. *Journal of Research in Science Teaching*, 36, 10 (1999b), 1085-1109.
15. Jordan, B., & Henderson, A. Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*, 1 (1995), 39-103.
16. Kerr, N. L., & Tindale, R. S. Group performance and decision making. *Annu. Rev. Psychol*., 55 (2004), 623-655.
17. Kozlowski, S. W. J., & Ilgen, D. R. Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7, 3 (2006), 77.
18. Landis, J. R., Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* 33 (1977), 159-174.
19. Larkin, J. H., and Simon, H. A. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11 (1987), 65-99.
20. Maxfield, M. G., & Babbie, E. R. Research methods for criminal justice and criminology: Wadsworth Pub Co (2010).
21. Norman, D. *The Design of Everyday Things*. New York: Doubleday, 1988.
22. Schafer, W.A., Ganoe, C.H. & Carroll, J.M. Supporting community emergency management through a geocollaboration software infrastructure. *Computer-Supported Cooperative Work: The Journal of Collaborative Computing*, 16 (2007), 501-537.
23. Stahl, G. *Group cognition*. Cambridge, MA: MIT Press, 2006.
24. Suthers, D. Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science*, 7, 3 (2001), 254-277.
25. Vygotsky, L. S. *Interaction between learning and development*. In M. Cole, V. John- Steiner, S. Scribner & E. Souberman (Eds.), Mind in Society: The Development of Higher Psychological Processes (pp. 79-91). Cambridge, MA: Harvard University Press, 1978.
26. Warner, N., Burkman, L., & Biron, C. H. Special operations reconnaissance (SOR) scenario: Intelligence analysis and mission planning, No. NAWCADPAX/TM-2008/184, 2008.
27. West, G. P. Collective Cognition: When Entrepreneurial Teams, Not Individuals, Make Decisions. *Entrepreneurship Theory and Practice*, 31,1 (2007), 77-102.
28. Woolley, A., Chabris, C., Pentland, A., Hashmi, M., & Malone, T. Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*, 6004 (2010), 686-688.