Impact of Roles on Participation and Task Performance

Susan Hahn, Michael Lewis, and Terri L. Lenox

School of Information Sciences, University of Pittsburgh
135 N. Bellefield Avenue, Pittsburgh, PA 15260 USA
(shahn@pitt.edu, ml@sis.pitt.edu, lenoxtl@westminster.edu)

Abstract

Effective team task performance requires that participants make appropriate use of all available knowledge and skills. In natural settings plagued by unreliable data and data sources, the difficulty of improving performance through aggregation of data increases with available resources. Teams must develop a shared perception of the task, knowledge of one another and of the reliability of each other's data to derive full value from pooled resources. Initial participant roles and expectations are refined to incorporate past experience and evolve interaction patterns recognizing individual strengths and weaknesses.

While our understanding of informational dynamics in human teams remains incomplete, even less is known about human-computer team dynamics. We hypothesized that effective human-computer performance also may require calibration of roles and expectations so that the decision maker can accurately interpret software behavior and anticipate its limitations. This calibration should incorporate expectations, experience and supporting evidence. We designed a target classification task in which a software agent played different roles varying from a simple aggregator of data through full-fledged decision maker. In experimental trials with 120 subjects, we found that higher roles of software participation positively influenced willingness to adopt aiding and resulted in improved task performance even under conditions of known errors.

Introduction

Effective performance in a situation involving more than one individual requires calibration of knowledge and decision-making capability among participants. The way each individual fits into the situation may be conceptualized as a role: a set of knowledge, experience, and interaction behaviors. Whether the work structure is peer group or hierarchical, the same elements apply to establishing roles: expectations, performance, feedback and adjustment.

Peer-to-peer groups are generally characterized as teams. A variety of team models have been developed by human factors researchers, e.g. Nieva, Fleishman and Rieck (1978); Cannon-Bowers and Salas (1997); Smith-Jentsch, Johnston and Payne (1998); Morgan, Glickman, Woodard, Blaiwes and Salas (1986); and Hinsz, Tindale & Vollrath, (1997). Most argue that in human teams, initial participant roles and behavior expectations are refined over time to incorporate past experience and evolve interaction patterns that recognize individual strengths and weaknesses.

The teamwork model described in Cannon-Bowers and Salas (1997) and Smith-Jentsch, Johnston and Payne (1998) consists of four dimensions: 1) Information exchange (exploit all available information sources; disseminate information; provide situation updates); 2) Supporting behavior; 3) Communication; and 4) Team initiative/leadership (provide feedback to team members; state clear and appropriate priorities). A basic tenet of this model is that teamwork skills exist independent of individual competencies.

Team performance, especially in tightly coupled tasks, is believed to be highly dependent on effective information processing. Members must understand how expertise is distributed and coordinate the management of this information (Wegner, 1987). Stasser and Titus (1987) found that groups in which expert roles were assigned reached correct decisions more often than groups who were not informed about the distribution of unique information. Since unshared information cannot be validated by others within the group, the contributing member must be able to convince the others that the recall is accurate and the others must feel confident in the source. It appears that expert role assignment helps to validate the unshared information for the group. Stewart and Stasser (1995) found that when expertise was assigned in collective recall and decision-making groups, more unshared information was discussed, and most of this information was mentioned in the written recall. On the other hand, in groups where roles were not assigned, less unshared information was discussed and very little of it was mentioned in the written recall. Team members need to be aware both of their own role and the roles of other experts within the group in order to take advantage of unshared information. Since others cannot validate unshared information, the contributing member must be able to convince the others that the recall is accurate and the others must feel confident in the source. It appears that expert role assignment helps to validate the unshared information for the group.

A model of groups as information processors from Hinsz et al. (1997) represents a shift in paradigm from individuals as information processors to groups as information processors. Groups process relevant and available information to perform intellectual tasks. This processing involves activities among the minds of group members and within the group. At the group level “information processing involves the degree to which information, ideas, or cognitive processes are shared, and are being shared, among the group members and how this sharing of information affects both individual- and group-level outcomes “ (Hinsz, Tindale & Vollrath, 1997, pg. 43). Failure to appropriately use team member capabilities may be caused by a mismatch of decision control and relevant knowledge (Sarter and Woods, 1994).

In hierarchical structures, job positions and titles form the basis for expectations regarding expertise, performance, and participation in decision-making. In addition, both supervisors and subordinates may exhibit individual preferences for participation. Both type of task and personal characteristics may interact to impact degree of participation (Abel-Halim, 1983). As in peer groups, the level of expertise among participants must be calibrated.

Performance may be judged by observing subordinate behavior or measuring outcome results. If the supervisor understands the behaviors necessary to succeed, and those behaviors are observable, feedback and calibration can occur at interim points during the task. In situations where the supervisor does not have extensive task knowledge, or the appropriate behavior is not observable, performance will tend to be judged on outcomes achieved (Ouchi, 1979). For complex tasks, supervisor knowledge of the task appears to be a key determinant in selecting outcome control (Kirsch, 1996). In that case, the subordinate will participate more in making decisions than in lower-level data processing.

Scully, Kirkpatrick and Locke (1995) found that in simulated supervisor/employee dyads higher task performance resulted when: 1) the supervisor had correct information and there was no participation from the employee; or 2) the employee participated and neither party had incorrect information or at least one party had correct information. Similar problems may arise in human-agent teams when a software agent has sole or preferential access to needed information.

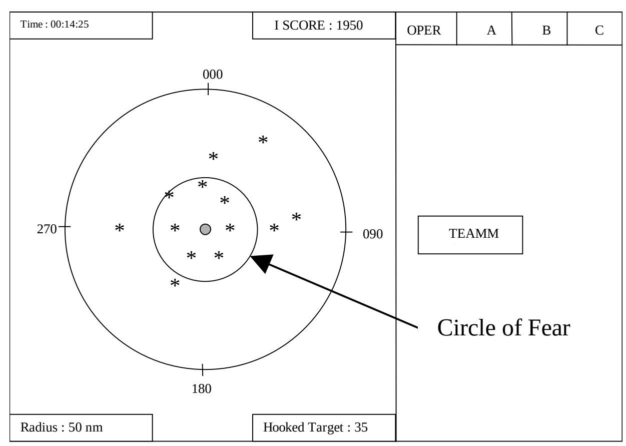
Appropriate use of information from others may depend upon experience, to calibrate the quality of data (Muir, 1987). However, in situations where people don’t have time or confidence in their own ability to complete a task, they are likely to rely on computer recommendations despite errors, because they believe that will lead to better overall performance (Lee and Moray, 1992).

Human-Human calibration often involves explicit discussion of expertise and information sources. In Human-Computer calibration the software may not be designed to explicitly assist in establishing capabilities and calibrating expectations or have flexibility in adaptation. We have conducted experiments that varied the role (information access, categorization, or decision) of software while holding the (potential) impact of the information it provided constant.

Method

We modified a moderate fidelity simulation (Tandem) of a radar target identification task, jointly developed at the Naval Air Warfare Center - Training Systems Division (Weaver, Morgan, and Hall, 1993) and the University of Central Florida to investigate human-computer interaction. During a simulation session sixty targets were presented, distributed in several concentric rings on the screen. The circle closest to the center is called the “circle of fear” and targets there should receive highest priority (Figure 1). Subjects ‘hooked’ a target by clicking on it, selecting ‘hook’ from the menu. They could then examine five parameters each to identify target *type* (air/surface/submarine), *source* (civilian/military) and *intent* (peaceful/hostile), and then decide on an appropriate response. The task imposed a high workload for information gathering and aggregation in a reactive, time critical situation.

Figure 1 - - The Tandem Display



We developed three roles for a software teammate, to assist in accessing and processing data for the decision task:

1. LIST (aggregated information) – an assembled list of parameters and values: climb/dive rate, speed, signal strength, and communication time;
2. TABLE (inferential information) -- a table showing parameter values classified by *type* in columns;
2. ORACLE (decision information) -- a message with an associated certainty factor e.g., "target 101is with .85 certainty a submarine". The raw parameter values remained available via an explanation button.

Subjects were required to use the menus alone for decisions on target *source* and *intent*; the teammate could be used to assist in classifying target *type* by accessing and processing four of the five relevant parameters. They were given a cue sheet for target type identification parameters (Table 1) and the remaining two tasks.

Table 1: Target Type Identification Parameters

	AIR	**SURFACE**	**SUB**
Current Speed	>35 knots	25-35 knots	<25 knots
Initial Altitude/Depth	> 0 feet	0 feet	< 0 feet
Initial Climb/ Dive Rate	> 0 feet/m	0 feet/m	<0 feet/m
Signal Strength	Medium	High	Low
Communication Time	0-40 secs	41-80 secs	81-120 secs

We also investigated effects of errors, which could occur in any role. In the control case, all three roles for aiding provided correct information so that data presented was derived directly from the Tandem menus. In other conditions errors were introduced in reading (data), categorization (classification) or in designating *type* (decision). Table 2 shows the agent roles and their potential types of errors. Data errors occurred when data displayed was different than data shown on the menus. This type of error was explained to the subjects as 'problems with the agent's sensors'. Classification errors occurred when data was placed into the wrong column in the table. For decision errors, type and certainty were assigned independent of the underlying parameters on error trials. Classification and Decision errors were explained to subjects as 'software problems'. Only one type of error was present for each subject, and the presence of each type of error was equated. For example, a decision-making teammate (Oracle) might commit data errors, classification errors, or decision errors; but each would support the same rate of erroneous decisions.

Table 2: Teammate Roles and Error Conditions

Errors	**LIST**	**TABLE**	**ORACLE**
None	X	X	X
Data	X	X	X
Classification		X	X
Decision Rule			X

For every target at least three of the data values agreed, to permit correct *type* identification. Error detection became increasingly difficult with greater roles of information aggregation. Data errors required comparing presented parameter values with those found on the menus. Classification errors required comparison between displayed category of a parameter and the proper categorization of its value. Detecting decision errors required performing the classification task manually and comparing the result.

Subjects were trained on the Tandem software for 10 minutes, followed by three 15-minute trials. Each subject saw one teammate role, with one of the four possible error conditions (Table 1). There were no "surprises" for role behavior or errors; actual teammate behavior in both training and trials was consistent with expected behavior. Subjects had free choice on the use of the teammate's information. They could request data processed by the agent or get raw data directly by accessing menus.

Independent variables were teammate role (list, table, oracle) and error condition. Dependent variables were task performance and reliance on the software teammate. Task performance was measured as: 1) number of targets hooked; 2) correct identification of target *type*, *source*, and *intent*; and 3) timeliness of target identification, i.e. time spent in the "circle of fear". We also recorded button presses to access the teammate or make menu selections, and several time measures.

Goals and Rational Decision Making

The goal for subjects was to maximize performance by getting information to correctly identify targets and either shoot or clear them, working from the *circle of fear* outward. Subjects could use the menus, the teammate, or both. However, due to the high workload conditions, subjects who relied on the menus alone would find it difficult to perform effectively. In the no-error condition, the decision to use a software teammate was a pure participation decision. In error conditions, participation should be modified by some verification of data. As mentioned above, data verification strategies varied with roles of data aggregation.

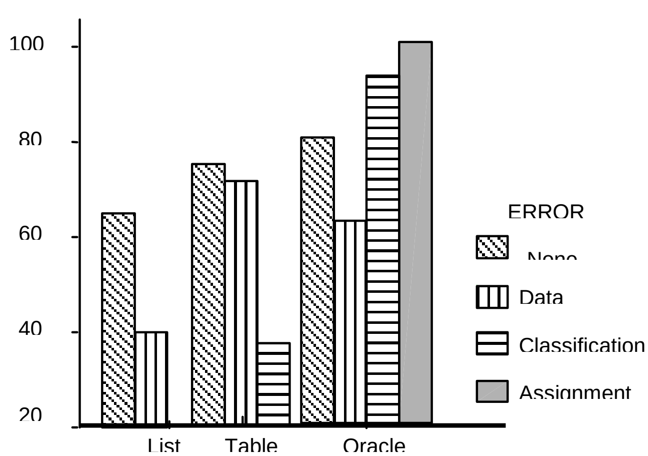
A minimum of 3 values will agree regardless of the error condition.

Level of Aggregation & Nature of Error	Rational Decision Maker Strategy (Minimum of 3 values will agree.)
Level 1 (List Error) - incorrect data values.	If at least 3/4 values agree, assign type classification based on values observed compared to cue-sheet values else (2/4) check fifth value from menu and assign type classification based on values observed compared to cue-sheet values.
Level 2 (Table Error) -- incorrect classification of data values into table (i.e., right data, wrong column).	If a comparison of aggregated data in table <> expected classification on cue-sheet, reclassify "misplaced" value into the correct column.
Level 2 (Data Error) – incorrect data values.	See strategy for Level 1 error.
Level 3 (Oracle Error) – incorrect conclusion drawn from correct data values	If confidence factor is greater than 70 % assign classification, else check values from table using the "see justification" button.

Results

Performance was analyzed using a repeated measures analysis of variance with session as the within subject factor and with software role of participation and error conditions as between group factors. Effects of session were significant ($p < .05$) for each of the dependent measures reported. Not surprisingly, subjects relied more on Role 1 (List) and Role 2 (Table) software teammates in the "no error" condition ($p < .01$). See Figure 2.

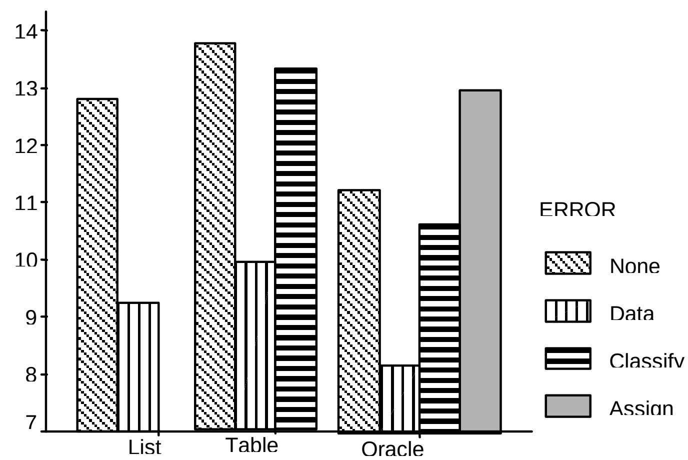
Figure 2: Number of Times Teammate Was Invoked



However, willingness to accept teammate participation for decision making from Role 3 (Oracle) was highest under all conditions ($p < .03$), despite the greater difficulty of error detection.

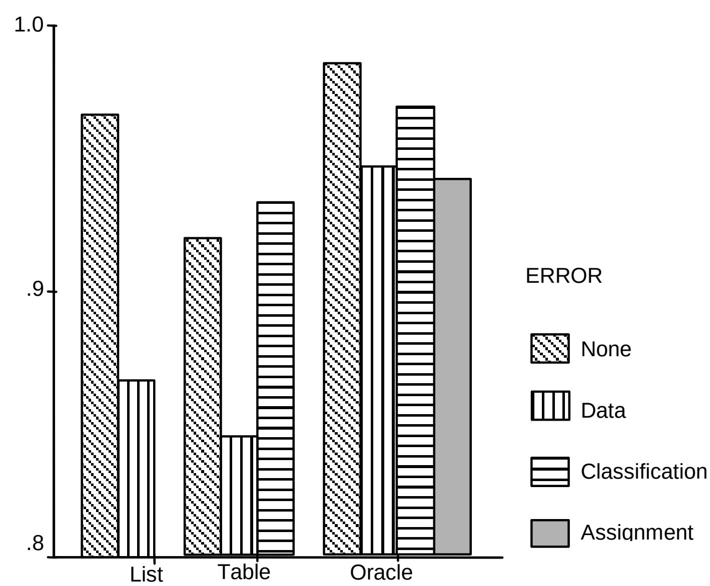
More targets were engaged in conditions without errors (see Figure 3).

Figure 3: Average Targets Hooked



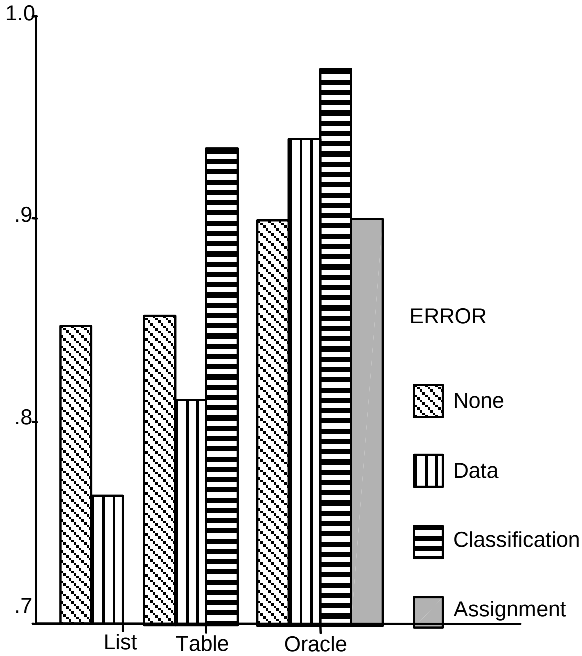
Both teammate role ($p < .02$) and error ($p < .01$) affected performance in the *type* identification task (Figure 4).

Figure 4: Percent of Targets Correctly Identified for Type of Target (Air/Surface/Submarine)



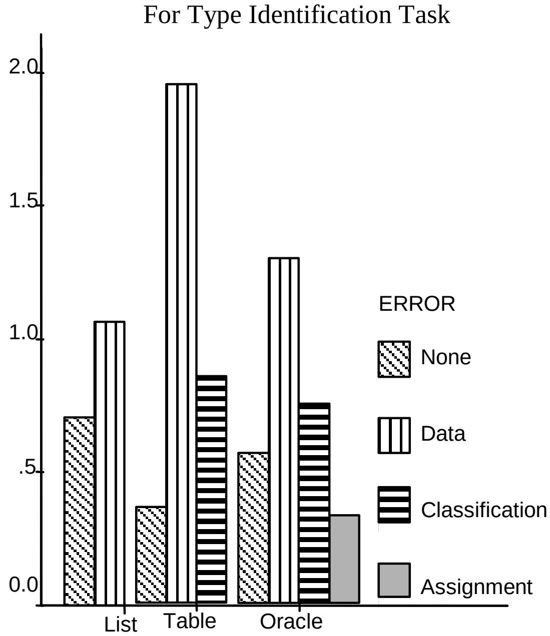
In addition, teammate role alone affected unaided portions of the task: correct civilian/military identification ($p < .05$) and correct shoot/clear decision ($p < .02$) (Figure 5).

Figure 5: Agent Type Alone Affected Civilian/Military Identification Task



Menus were accessed more per target for when data errors occurred ($p < .001$) than when classification or assignment errors or no errors occurred (Figure 6).

Figure 6: Average Number of Menu Accesses Per Target For Type Identification Task



Discussion

The experimental task had speed v. accuracy tradeoffs to maximize performance. Since subjects could not reasonably complete the task manually, use of the software teammate appeared to be attractive despite the presence of errors. Performance was not penalized for "guessing", therefore one successful strategy was to use the software teammate even if the number of expected errors was high. Expected and actual behavior of the teammate was consistent for both role of participation and quality of information. This situation is analogous to participation in a human team where overall performance can be enhanced by shared effort, with occasional errors offset by the total volume of work completed. We would expect to see different tolerance in a zero-error or zero-defect task.

The cognitive efficiencies offered by the Oracle, compared to lower roles of participation in decision making, appeared to outweigh the effects of unobservable errors. While subjects in other conditions performed no better than their models, Oracle users performed better at both aided and unaided tasks.

Why did people use the Oracle? The role was to provide a decision for the user that could be accepted or rejected. The users did not have to check the cue sheet or the menu, which reduced their cognitive workload and allowed them to focus on the non-aided tasks. Therefore, use of Oracle was easy. It was possible to check the reliability using the "see justification" button for decision rule and classification based errors. However, data errors were hidden. In this situation, time pressure likely motivated subjects to tolerate occasional errors in order to complete a greater proportion of the task and maximize overall performance.

Consistent with Muir (87), some subjects elected not to use any software teammate when told it might not be reliable. Such a "non-adoption" might be influenced by subjects' preferences for retaining control over decision-making.

Acknowledgement

This research was supported by an Office of Naval Research grant N-00014-96-1-1222.

References

Abdel-Halim, A.A. (1983). Effects of task and personality characteristics on subordinate responses to participative decision making. *Academy of Management Journal* 26(3): 477-484.

Cannon-Bowers, J.A. and Salas, E. 1997. A framework for developing team performance measures in training. In M.T. Brannick, E. Salas, & C. Prince (Eds.). *Team performance assessment and measurement: Theory, methods and applications*. Hillsdale, NJ: Erlbaum, 45-62.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin* 121(1):43-64.

Kirsch, L.J. (1996). The management of complex tasks in organizations: controlling the systems development process. *Organization Science* 7(1):1-21.

Lee, J. and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35:1243-1270.

Morgan, B. B., Jr., Glickman, A. S., Woodard, E. A., Blaiwes, A. & Salas, E. (1986). Measurement of team behaviors in a Navy environment (NTSC Report, No. 86-014). Orlando, FL: Naval Training Systems Center.

Muir, B.M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies* 27:527-539.

Nieva, V. F., Fleishman, E. A. & Rieck, A. (1978). Team dimensions: Their identity, their measurement and their relationships. Contract No. DAH19-78-C-0001, Response Analysis Corporation, Washington, DC.

Ouchi, W.G. (1979) A conceptual framework for the design of organizational control mechanisms. *Management Science* 25(9):833-848.

Sarter, N. B. and Woods, D. D. 1994. Pilot Interaction with Cockpit Automation II : An Experimental Study of Pilot's Model and Awareness of the Flight Management System (FMS). *International Journal of Aviation Psychology* 4(1):1-48.

Scully, J.A., Kirkpatrick. S.A. and Locke, E.A. 1995. Locus of Knowledge as a Determinant of the Effects of Participation on Performance, Affect, and Perceptions. *Organizational Behavior and Human Decision Process* 61(3):276-288.

Smith-Jentsch, K. Johnston, J.H., and Payne, S.C. 1998. Measuring team-related expertise in complex environments". In J. A. Cannon-Bowers & E. Salas (Eds.). *Decision Making Under Stress: Implications for Individual and Team Training*. Washington, DC: American Psychological Association, 61-88.

Stewart, G., and Stasser, D.D. (1995). Experet roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology* 31:244-265.

Stasser, G., and Titus, W. 1987. Effects of information load and percentage of shared information on the disseminationof unshared information during group discussion. *Journal of Personality and Social Psychology* 53:81-93.

Weaver, J.L., Morgan, Jr., B.B. and Hall, J. 1993. Team Decision Making in the Command Information Center: Development of a Low-Fidelity Team Decision Making Task for Assessing the Effects of Teamwork Stressors. Naval Training Systems Center, Technical Report 92-xxx[sic], February 1993.

Wegner, D. 1987. Transactive memory: A contemporary analysis of group mind. In B. Mullen and G.R. Goethans (Eds.). *Theories of group behavior*. New York: Springer-Verlag, 185-208.