*Theoretical Issues in Ergonomics Science*
Vol. 6, No. 2, March–April 2005, 157–171

Taylor & Francis
Taylor & Francis Group

# A theoretical framework and quantitative architecture to assess team task complexity in dynamic environments

L. ROTHROCK†*, C. M. HARVEY‡ and J. BURNS§

† Pennsylvania State University, 310 Leonhard Building, University Park, PA 16802, USA
‡ 3128 CEBA, Louisiana State University, Baton Rouge, LA 70803, USA
§ Sonalysts, Inc. Suite 260, 12501 Research Parkway, Orlando, FL 32826, USA

The predominance of team decision-making and performance assessment literature has defined team measures as highly abstract concepts (e.g., team leadership, competence, innovation, empowerment). Likewise, a clear taxonomy for defining team tasks has remained elusive. Thus, this paper presents a framework by which to classify team tasks based on two basic premises: (1) a team task can be broken down into quantifiable components; and, (2) team performance can be used to evaluate a task's complexity relative to another task. This framework relies on the ability to objectively measure individual team member subtasks relative to a team objective that is composed of several windows of opportunity that must be achieved by individual members to achieve good team performance. This proposed theoretical framework takes a simulation-based approach by which to evaluate team tasks and performance. The approach is driven by the need to understand team tasks and their relative performance in military, government, and commercial applications.

*Keywords:* Team decision-making; Task complexity; Team performance

## 1. Introduction

On the morning of February 14 1982, the 84-member crew of the drill rig *Ocean Ranger* began to prepare for the coming storm in the North Atlantic (based on the account of the incident by Chiles (2001)). Although the rig floated much like a pontoon boat and supported a massive infrastructure weighing up to 14,000 tons, the crew members were sheltered in their quarters and confident that the rig would be able to withstand the 55 ft waves. In fact, the rig was built to survive hurricane weather of 110 ft waves. At approximately 8 pm, however, ocean waves break a small portlight window to the ballast control room located near sea level and doused the electronics with seawater. Because seawater is a conductor, the shorted circuits in the room began to cause ballast tanks to flood and empty by themselves. After a few minutes, the activities in the ballast tanks began to cause the rig to list beyond the prescribed limits. About an hour later, the rig's electrician cut off the power to the control room, thereby closing all tank valves and preventing further flooding. While the rig appeared to stabilize, it was still below its 'survival draft' to better

*Corresponding author. Email: lrothroc@psu.edu

*Theoretical Issues in Ergonomics Science*
ISSN 1463–922X print/ISSN 1464–536X online 
http://www.tandf.co.uk/journals
DOI 10.1080/1463922042000295678

survive storm waves. The ballast-control operator then made the fatal decision to turn on the power and attempt to raise the rig to above survival draft. Unaware of the impact of continuing electrical shorts on the state of the ballast tanks and unfamiliar with the emergency operations procedures, the operator was unable to right the ship. The rig eventually listed to such a degree that water began pouring into the rig through external storage compartments. At around 3 am, the *Ocean Ranger* disappeared from view. There were no survivors.

On the morning of May 17 1987, an Iraqi fighter jet approached within 25 nautical miles of the frigate *U.S.S. Stark*. The fighter used its *Cyrano IV* radar to target the frigate, fired two *Exocet* missiles, and returned to an Iraqi airbase. Personnel onboard the *Stark* had been tracking the fighter for over an hour and, upon detection of the *Cyrano IV* emission, issued warnings to the fighter. Because the U.S. was not at war with Iraq, there was no reason for the crew to suspect an attack. Nevertheless, four minutes later, both *Exocets* struck the *Stark*, killing 37 of the 221 crew onboard. No defensive weapons were ever deployed by the *Stark*.

The two events are provided to illustrate differences that exist between teams in contrasting task environments. There is little debate that it is not possible to reproduce the situation faced by either crew in a laboratory setting. Nevertheless, in an effort to characterize teams in operational contexts, the predominance of team decision-making and performance assessment literature has defined team measures as highly abstract concepts. Terms such as team leadership, competence, innovation, and empowerment are replete in team literature (Brannick *et al.* 1997, Smith-Jentsch *et al.* 1998). While these terms are intuitively appealing (e.g., successful teams are empowered), they are often without specified mathematical meanings and, therefore, without quantifiable relationships to other team constructs. Characterizations that most successful teams are well-led, innovative, competent, and empowered are helpful criteria for expert practitioners who have their own sense of what these terms mean. However, they are less useful when applied toward the design of training systems for future team technologies. In contrast, we propose a quantitative synthetic architecture to objectively assess team performance in a dynamic decision making environment.

### 1.1 *Background*

Most complex decisions involve many data, human, and technological sources collaborating to support decision makers. However, when responsibility for task accomplishment moves from the province of one person to a multitude of natural and artificial intelligences, the system changes quantitatively and qualitatively. Quantitatively, the system is more complex and dynamic. This complexity increases further as the constituent intelligences are separated in time. Qualitatively, the system exhibits properties that were not evident when a lone individual is working on a set of tasks.

Thus, the tools and methodologies that have been developed to understand the work of the sole individual do not necessarily accommodate the interaction of multiple members. Much work has been done to attempt to understand the team environment in the form of cognitive engineering, computer cooperative supported work, and groupware. Much of the work in this area focuses on teams that create a common artifact where debate and negotiation are often not constrained by time. While this research has provided some insight into how teams operate, in

Table 1. Inconclusive team research.

| | |
|---|---|
| **Time to complete task** | |
| Distributed took longer (Weeks and Chapanis 1976, Hiltz *et al.* 1986) | No difference (Bul and Slvasankaran 1990) |
| **Satisfaction** | |
| Distributed less satisfied (Gallupe *et al.* 1988) | No difference (Bul and Slvasankaran 1990) |
| **Alternatives** | |
| Distributed—more alternatives (Lewis 1982, Dennis *et al.* 1990) | No difference (Hammond 1998) |
| **Consensus** | |
| Distributed—less agreement (Hiltz *et al.* 1986, Williams 1977) | No difference (Watson *et al.* 1988) |

many cases it has been inconclusive on elements that impact team performance (e.g., table 1).

In contrast, supervisory control teams engage in a much more dynamic and event driven situation where negotiation and debate are not generally feasible. These environments, instead, require the timely coordination of activities (Jasek and Jones 2001). In fact, much has been done in the attempt to develop team performance measures (cf. Fleishman and Zaccaro 1992, Cannon-Bowers and Salas 1997a, b, Cannon-Bowers and Salas 1998). The bulk of this work has focused on existing man–machine systems where automation, interfaces, and configuration (to name a few factors) were seen as fixed constraints.

## 2. Premise for quantifying team performance

The predominance of team decision-making and performance assessment literature has defined team measures as highly abstract concepts (e.g., team leadership, competence, innovation, empowerment) (Brannick *et al.* 1997, Smith-Jentsch *et al.* 1998). While these concepts are not without merit, most practitioners and non-practitioners alike are likely to maintain their own unique definition of what exactly these terms mean in implementation. As a result, they are of less utility when informing the design of training systems for future team technologies.

Likewise, we submit that the reason that researchers find different outcomes for similar tasks is largely based on their definition of the task. One researcher may call his task complex. Similarly, another researcher may call her task complex. Yet the very nature by which they define complexity is different. Thus, the field of team performance assessment is inhibited by the ability to compare scientific studies. One may contend that theories can only be assessed when a field has common agreeable measurement methods. Thus, the development of new team theories is inhibited by the ability to classify team tasks.

To begin to answer the question of what constitutes team task complexity as well as to provide researchers and trainers with information they can use to create improved man-machine interfaces, we propose that team research must undertake a systematic approach. Not only must we understand how the human collaborates—whether with another human or machine—we must also be able to characterize how a particular implementation to support collaboration (e.g., technology or process) impacts the team collaborative performance. We propose a framework for decision-making assessment based on findings in simulation-based experimentation.

The use of simulations in the study of complex decision making behavior and team interactions has yielded success in recent years (Bondanella *et al.* 1999, Handley *et al.* 1999, Sauer *et al.* 2000). However, this success is tempered by the need for situation-specific modeling (Pew and Mavor 1998). Specifically, this paper proposes a rich synthetic task environment (STE) that allows for a diverse range of empirical investigations into current and emerging man-machine systems and would provide a means for classifying team tasks along a common yardstick.

The approach we will discuss is driven by the need to understand team performance in military, government, and commercial applications. We believe this research will have fundamental impacts in areas such as air traffic control, emergency operations, and nuclear plant operations.

While our paper is predicated on the need to classify team tasks, our STE environment would allow researchers and practitioners to answer questions such as:

- How can we accomplish a given task with less people?
- What skills are needed to accomplish the task?
- How should the work task be organized to facilitate collaboration between operators?
- What training will result in the greatest team performance?
- Does changing the technology impact individual/team performance?

### 2.1 *Definition of task environments*

Our starting point is the concept of team task complexity proposed by Harvey (1997, 2001). The concept is a synthesis of existing literature (Daft and Macintosh 1981, Wood 1986, Campbell 1988, Campbell 1991, Byström and Järvelin 1995) and defines complexity along three primary characteristics: scope, structurability, and uncertainty. Since the metric forms the foundation by which teams collaborate, the quantification of tasks is essential to allow researchers to compare experimental results. Table 2 shows the team characteristics along the dimensions proposed by Harvey (1997, 2001).

The task scope is the breadth, extent, range, reach, or general size of a task. The scope is a function of the sub-tasks, outcome(s), information processed, and the outcome characteristics and their conflicting objectives. Each task can be decomposed into sub-tasks. A sub-task has identifiable behaviors or steps with an identifiable purpose or direction. Outcomes are entities that result from activities of the collaborative individuals and are independent of the behaviors used to produce them. For each outcome, there exists a set of characteristics by which its success is measured. Outcome characteristics include the attribute, aspect, property, quality, or trait. Characteristics may conflict with each other and thus increase the complexity of the task. For example, altitude and accuracy may conflict with each other in an aerial intelligence seeking information task. The last element that defines the task scope is information. Information is the amount of required knowledge in the accomplishment of the task.

With this basic understanding of a task and its scope, the other two dimensions can be explained. Task structurability represents how well-defined the sequence and

Table 2. Team characteristics.

| Characteristic | Metric |
|---|---|
| **Scope** | |
| Sub-tasks | Number of sub-tasks available |
| Outcomes | Number of possible outcomes |
| Outcome Characteristics | Number of ways success is measured |
| Characteristic Conflicts | Number of competing objectives |
| Information | Number of variables required to provide the necessary information to generate outcomes |
| **Structurability** | |
| Analyzability | Number of sub-tasks with imperfect mappings to outcomes |
| Alternatives | Number of paths available to reach desired outcome characteristics |
| Coordination | Number of required relationships among sub-tasks |
| **Uncertainty** | |
| Internal confidence | Number of imperfect mappings among task alternatives, sub-tasks, and characteristics |
| External constraints | Number of real-time changes in the set of required outcome characteristics |
| Random events | Expectation of number of chance occurrences or irregular events during course of a task |

relationships between subtasks are, and are determined by the elements analyzability, alternatives, and coordination. Analyzability reflects the degree of consistency between sub-tasks and their outcomes. If characteristics reflected by an outcome can be reached in more ways than one, the number of paths to reach it is summed as task alternatives. Moreover, if task accomplishment is contingent on coordination among sub-tasks, the number of relationships required is counted as task coordination.

The task uncertainty dimension measures complexity based on the degree of predictability or confidence associated with a task. Internal confidence indicates the degree of certainty or predictability of the structure established among tasks, alternatives, sub-tasks, and characteristics. External events include changes in the set of required product characteristics that are imposed by higher echelons of command. It is worth noting that random events have been included since these chance events can ultimately affect a task's complexity.

Using our task characteristics, we suggest that a three-dimensional *team complexity space* (see figure 1) exists where vastly different team environments can be placed.

We will return to our proposed utilization of the team complexity space shortly. For now, we turn our focus to the development of a team task performance measurement system based on a temporal extension of signal detection theory known as time windows.

### 2.2 *Definition of individual and team task performance*

A prerequisite for designers of complex systems is a proper understanding of operator performance characteristics. While human factors texts provide some insights into basic performance issues, the emergence of highly automated computing systems have fundamentally altered the way humans work. Our approach to quantify and analyze human performance within a complex, time-critical system is centered on a measurement construct, called a time window, which enables

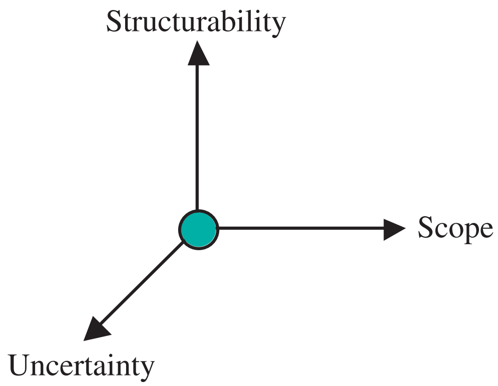


Figure 1. Task Complexity Space.

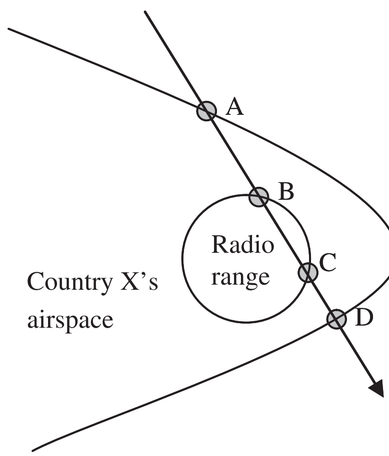


Figure 2. Air Traffic Control example. An aircraft enters country X's airspace at Point A, enters range to establish radio contact at Point B, leaves radio range at Point C, and leaves country X's airspace at Point D.

a functional relationship between constraints on team activities and time availability (Rothrock 2001).

To illustrate time windows, consider the following example: an air traffic control (ATC) center monitors aircraft entering and leaving the airspace of country X (figure 2). Assume that the center is responsible for the identification of all unknown aircraft entering the airspace, establishing radio contact with all aircraft that come within radio range, and providing emergency assistance services to aircraft in distress. A private civilian aircraft, traveling along the trajectory indicated by the direction vector, enters country X's airspace at point A, enters range to establish radio contact at point B, reports a medical emergency onboard the aircraft and leaves radio range at point C, and leaves country X's airspace at point D.

Consider an ATC center that consists of two people—a supervisor and a radio operator. Assume that the supervisor has been given specific instructions to electronically mark the identity of all unknown aircraft entering the airspace. To assist the supervisor, the radio operator uses the one radio to make contact with aircraft within range. Both people are responsible for reporting emergencies to the local emergency services dispatcher via a shared telephone.

We define a time window as a construct that specifies a functional relationship between a required situation and a time interval that specifies availability for action. A time window does not specify what action must be taken, but only that there exists an action that will result in the required situation. Moreover, in a team context, we assume there exists shared time windows among all combinations of team members. For example, in the ATC center introduced earlier, there are three types of windows—one for the supervisor (to identify aircraft), one for the radio operator (to establish radio contact with aircraft), and one shared between the team members (to report emergencies to the dispatcher). Mathematically, we define the number of potential time windows as $\sum_{r=1}^{m} m!/(r!(m-r)!)$ for a team with $m$ members. That is, there are potentially shared windows among all possible team sub-groups. We therefore expand the notion of time window presented in Rothrock (2001) to account for a team context so that for $m$ team members, there exists $w_{i,j}$ time windows for $i = 1$ to $\sum_{r=1}^{m} m!/(r!(m-r)!)$ and $j = 1$ to $n_i$ where $n_i$ is the number of windows within each window type.

At the onset of team interaction, all time windows are designated as inactive and represented by the set $U_0$. Until a time window is designated as open, it remains inactive. Time windows are designated as open if the availability for action exists for a required situation at the current point in time space. The set of open time windows at time $t$ is designated as $O_t$. When a required situation no longer exists, the corresponding time window is designated as closed. The set of closed time windows at time $t$ is denoted as $C_t$. The membership of $U$, $O$, and $C$ is defined to be persistent over time (i.e., $U_{t+1} = U_t$, $O_{t+1} = O_t$, and $C_{t+1} = C_t$) unless designated otherwise. Conditions specifying the opening and closing of time windows are dependent on physical laws of nature as well as rules and regulations of the task domain.

To complete our construct in a temporal context, one must define operator action and the relationship between action and time window. An operator action is defined here as a two-tuple that includes an act performed by the operator at a specific point in time. In the course of operator interaction within a dynamic task environment, $n$ actions are denoted as $\mathbf{b}_h$ for $h = 1$ to $n$. The relationship between action and time window can be described by two Boolean indicator functions, $I_w^l$, such that, for $l = 1$, the function evaluates whether an action meets the required situation specified by a time window, and for $l = 2$, the function evaluates the relevance of an action toward a time window.

Thus,

$$I_w^1(\mathbf{b}) = \left\{ \begin{array}{ll} 1 & \text{if } \mathbf{b} \text{ meets situation specified in } w \\ 0 & \text{if } \mathbf{b} \text{ does not meet situation} \end{array} \right\}, \text{ and}$$

$$I_w^2(\mathbf{b}) = \left\{ \begin{array}{ll} 1 & \text{if } \mathbf{b} \text{ is relevant toward } w \\ 0 & \text{if } \mathbf{b} \text{ is not relevant toward } w \end{array} \right\}$$

We construct six predicates, $M_T^k(w, \mathbf{b}_h)$ for $k = 1$ to 6, to characterize fundamental relationships between time windows and operators actions over a time interval $T$. In particular, the truth value, $\|M^k(w, \mathbf{b}_j)\|_{T+, T-}$, of each predicate is evaluated for a time interval that starts when team interaction in the task begins ($T+$) and ends when team interaction ceases ($T-$). Given that $\mathbf{b}_h$ occurs at time $t$, equations to evaluate the first five predicates are listed as follows:

- An on-time action that results in a required situation, $M^1_T(w_{i,j}, \mathbf{b}_h)$, is defined as:

$$\|M^1(w_{i,j}, \mathbf{b}_h)\|_{T+,T-} = 1 \quad \text{iff } \exists\ i,j \text{ such that } [I^1_{w_{i,j}}(\mathbf{b}_h) = 1] \wedge (w_{i,j} \in O_t); \tag{1}$$

- An early action that results in a required situation, $M^2_T(w_{i,j}, \mathbf{b}_h)$, is defined as:

$$\|M^2(w_{i,j}, \mathbf{b}_h)\|_{T+,T-} = 1 \quad \text{iff } \exists\ i,j \text{ such that } [I^1_{w_{i,j}}(\mathbf{b}_h) = 1] \wedge (w_{i,j} \in U_t); \tag{2}$$

- A late action that results in a required situation, $M^3_T(w_{i,j}, \mathbf{b}_h)$, is defined as:

$$\|M^3(w_{i,j}, \mathbf{b}_h)\|_{T+,T-} = 1 \quad \text{iff } \exists\ i,j \text{ such that } [I^1_{w_{i,j}}(\mathbf{b}_h) = 1] \wedge (w_{i,j} \in C_t); \tag{3}$$

- An action that is relevant toward a required situation, but does not result in it, $M^4_T(w_{i,j}, \mathbf{b}_h)$, is defined as:

$$\|M^4(w_{i,j}, \mathbf{b}_h)\|_{T+,T-} = 1 \quad \text{iff } \exists\ i,j \text{ such that } [I^1_{w_{i,j}}(\mathbf{b}_h) = 0] \wedge [I^2_{w_{i,j}}(\mathbf{b}_h) = 1]; \tag{4}$$

- An action with no corresponding time window, $M^5_T(\mathbf{b}_h)$, is defined as:

$$\|M^5(\mathbf{b}_h)\|_{T+,T-} = 1 \quad \text{iff } \forall\ i,j,\ (I^2_{w_{i,j}}(\mathbf{b}_h) = 0). \tag{5}$$

Because the sixth predicate is based on a time window instead of action, the equation to evaluate it is defined separately as follows:

- A time window, $w_{i,j}$, that has been missed, $M^6_T(w_{i,j})$, is defined as:

$$\|M^6(w_{i,j})\|_{T+,T-} = 1 \quad \text{iff } \forall\ j,\ (I^2_{w_{i,j}}(\mathbf{b}_h) = 0). \tag{6}$$

Because of their reliance on temporal logic, equations (1–6) offer a framework that can now be utilized as a dependency relation between an action and a required situation that is also bound by time. Time windows represent a belief system of required situations that are time and environment-based. Therefore, a truth maintenance system (TMS) (Doyle 1979) is needed to maintain time windows throughout the timeframe of team interaction. Instead of presenting the details of the truth maintenance system implementations, this paper focuses on the time windows framework and outcomes. For examples of TMS use in human–machine interaction research, see Rothrock (2001) and Furuta and Kondo (1993).

The utility of a time window is not only in its temporal and functional descriptions, but also in the richness of the possible outcomes. Some time window outcomes have already been described. The complete space of possible time window outcomes (see figure 3) is represented by the fundamental relationships between time windows and operator actions outlined in equations 1–6. In itself, the existence of a required situation does not impact team performance. It is the presence of operator action in a temporal context that specifies whether performance is good or poor. An incorrect action is represented as equation (4). A correct action, on the other hand, can be further characterized as early (equation (2)), on-time (equation (1)), or late

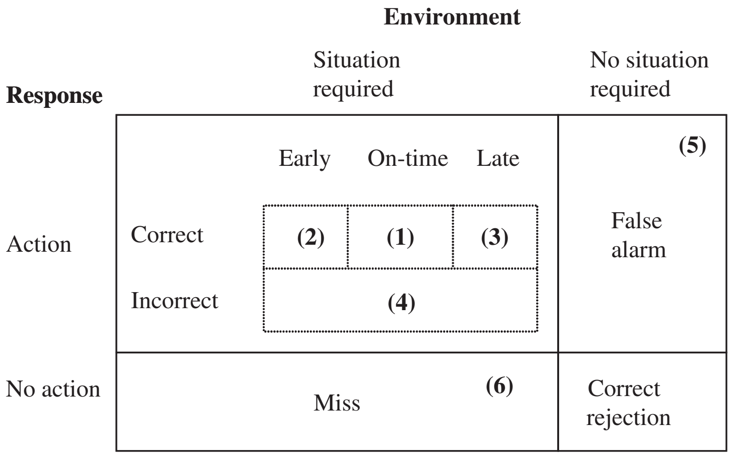


Figure 3. Possible time window outcomes. The environment is delineated in terms of situation required (time window exists) or no situation is required (time window does not exist). Equations (1–4) represent actions that are relevant to a time window. Equations (1–3) represent actions that result in the required situation (correct actions). Equation (4) represents actions that do not meet the required situation (incorrect actions) even though they are relevant.

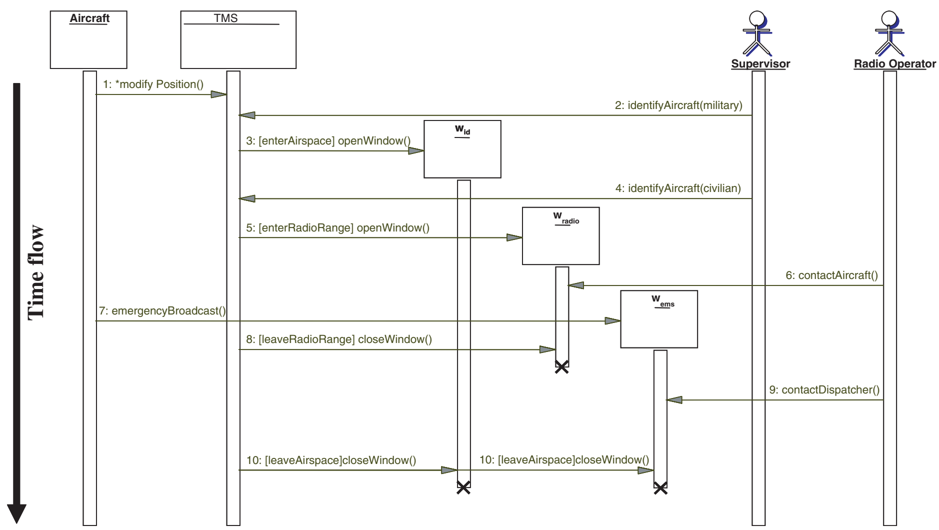


Figure 4. Sequence diagram of ATC scenario.

(equation (3)). An action with no corresponding required situation is categorized as equation (5). A non-action for an existing situation requirement is characterized as a miss and is represented as equation (6).

Revisiting the ATC center example, the event flow of operator actions and aircraft movements is reflected in the sequence diagram shown in figure 4. A sequence diagram (Booch *et al.* 1999) is a dynamic representation of a system which focuses on event-driven state changes. Objects that effect or accept change in the ATC example include the aircraft, the truth maintenance system (TMS), the supervisor, and the radio operator. Each action event originates from an object and is passed to one or more objects in the system. The flow of time is indicated as the relative ordinate position of the event. A chronologically-ordered narration on the sequence of events follows:

1. The flight of the unknown aircraft along the southeasterly trajectory is accomplished by the iterative call of the modifyPosition() method.
2. The first supervisor action, $\mathbf{b}_1$, of incorrectly identifying the aircraft (as a military aircraft) is posted to the TMS. However, this is an incorrect identification action that has been taken early. The window with a required situation that matches the identification action is the first identification window, $w_{id,1}$, so that the predicate satisfied is $\|M^4(w_{id,1}, \mathbf{b}_1)_{T+,T-}\|$.
3. When the TMS detects the aircraft entering the airspace of country X, the identification window, $w_{id,1}$, is designated open.
4. The second supervisor action, $\mathbf{b}_2$, correctly identifies the aircraft (as a civilian aircraft) and is posted to the TMS. Since the action is correct and on-time, the predicate satisfied is $\|M^1(w_{id,1}, \mathbf{b}_2)_{T+,T-}\|$.
5. When the control detects the aircraft entering radio range, the radio contact window, $w_{\mathrm{radio},1}$, is designated open.
6. The first radio operator action, $\mathbf{b}_3$, is to contact the aircraft after it comes within range. This action is correct and on-time. Therefore, the predicate satisfied is $\|M^1(w_{\mathrm{radio},1}, \mathbf{b}_3)_{T+,T-}\|$.
7. After radio contact is established, the aircraft broadcasts an emergency distress message, which opens the window to report the emergency to the dispatcher, $w_{ems,1}$. Unlike the identification window and the radio contact windows, which are designated for individual members, this window is shared by the team. Therefore, either the supervisor or the radio operator can take the action.
8. When the TMS detects that the aircraft is no longer within radio range, $w_{\mathrm{radio},1}$ is closed.
9. Before the aircraft leaves Country X, the radio operator contacts the dispatcher to report the emergency. This action, $\mathbf{b}_4$, is taken correctly before the window closes, and satisfies $\|M^1(w_{ems,1}, \mathbf{b}_4)_{T+,T-}\|$.
10. When the TMS detects the aircraft leaving country X's airspace, both $w_{id,1}$ and $w_{ems,1}$ are designated closed.

In our example, the team performed one early incorrect action and three on-time correct actions. More importantly, the TMS dynamically determines opportunities for action so that stated objectives of the ATC center can be met.

The normative course of scientific investigations is to narrow the scope of a real-world task into laboratory tasks that are generalizable. The emphasis, therefore, is on finding methods that analyze factors in isolation. However, Hammond (1986) notes that research on dynamic and complex environments should take place in representative settings because findings from simple laboratory forced-choice tasks cannot readily be transferred to operational environments. Recognizing the problem, researchers have sought to develop techniques to measure performance in individual user tasks that are more representative of the operational environment (Sanderson 1993, Howie and Vicente 1998, Raby and Wickens 1994, Laudeman and Palmer 1995).

Even with a focus on the environment, however, the majority of team measurement efforts remain guided strictly by operator actions, and do not adequately consider the environment in which the task is situated (Cannon-Bowers and Salas 1997b, 1998). The time window construct represents a fundamental shift from existing performance measurement approaches. It provides a computational framework to dynamically evaluate heterogeneous situation demands and the abilities of the team members to meet them in a complex domain.

### 2.3 *Establishing team profiles in team complexity space using time windows*

We propose that, by combining the framework to specify task complexity in the team complexity space with the use of time windows, profiles of the way teams interact given different task environments can be established. In order to establish these profiles, we must first determine the feasibility of regions of the team task complexity space to enable implementation of performance measurement systems. Specifically, at each point in the team complexity space, there exists a task structure that specifies:

- the realizability of human-in-the-loop simulations;
- the constructability of time window measures;
- and the implementability of interaction measures.

Therefore, we do not presume that all team tasks can be fully simulated and measured—the incident with drill rig *Ocean Ranger* is a prime example. Nevertheless, we submit that there are regions in the complexity space that do support simulation and measurement. Moreover, by sampling the space with different tasks (e.g., figure 5), we suggest that one can build different team performance profiles that characterize individual and team responses as well as team interaction.

### 2.4 *Comparing complexity space profiles*

Consider the ATC task characteristics presented in table 3. The ATC example provides one data point in the complexity space. The task is quite simple,

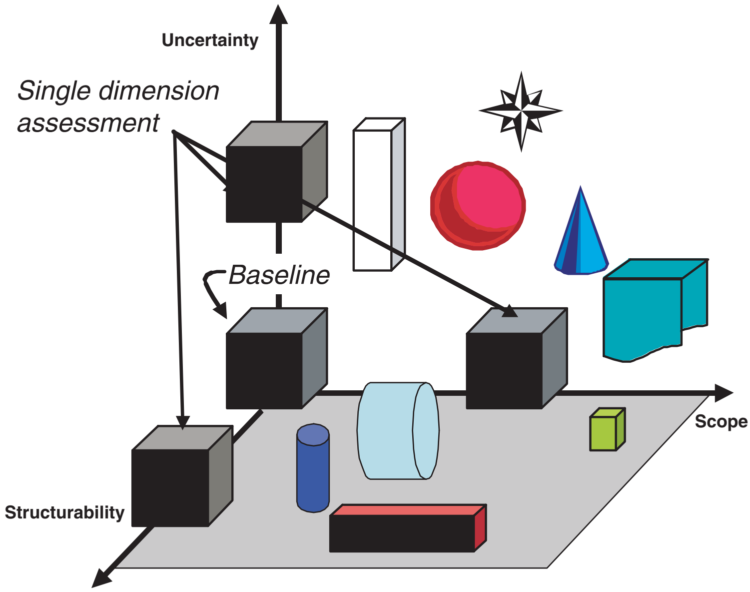


Figure 5. Team task complexity profiles.

Table 3. Team characteristics for ATC example.

| Characteristic | ATC Example Values |
|---|---|
| **Scope** | |
| Sub-tasks | 1. identify all aircraft entering airspace<br>2. establish radio contact<br>3. determine if aircraft is in distress<br>4. provide emergency assistance service |
| Outcomes | 1. aircraft identity type<br>2. radio contact established<br>3. emergency assistance provided (for each aircraft) |
| Outcome characteristics | 1. compare true aircraft identity<br>2. success to establish radio contact<br>3. success to provide emergency assistance |
| Characteristic Conflicts | No competing objectives |
| Information | 1. radar contact data<br>2. radio transmissions from aircraft |
| **Structurability** | |
| Analyzability | No sub-tasks with imperfect mappings to outcomes |
| Alternatives | Multiple paths available to reach desirable outcome |
| Coordination | 1. identification is dependent on the proximity (radar contact data) of the aircraft<br>2. the need for emergency assistance is dependent on first establishing radio contact |
| **Uncertainty** | |
| Internal confidence | No imperfect mappings |
| External constraints | No real-time changes in the set of required outcomes |
| Random events | No expectation of change events during the course of task execution. |

in comparison to both the *Ocean Ranger* or *Stark* incidents. There are few sub-tasks (4), few possible outcomes for each aircraft (3), no competing objectives, no imperfect mappings to task outcomes, and low uncertainty.

The Task Complexity Space (figures 1 and 5) contains an origin and implies a ratio scale. However, comparisons between profiles in the complexity space are proposed along an ordinal scale initially. That is, while different tasks can be rank ordered, the functional relationship along each dimension (e.g., additive, multiplicative) must be determined empirically. We submit that a ratio scale will eventually be established. However, this endeavor will require a collaborative research enterprise that spans disciplines and universities to share and compare results along the spectrum of the complexity space.

## 3. Discussion and conclusion

To evaluate the complexity space using the proposed theoretical framework, a synthetic task environment is being developed to emulate dynamic team tasks. While there does not exist a formal definition of a STE, there exists agreement among researchers that synthetic task environments lie between simple lab tasks and the outside world (Gray 2002). For our purposes, we define a STE as a human-in-the-loop simulation-based environment used to assess team performance as it pertains to function allocation, human interface design, and selection of decision support methods. A configurable STE architecture will enable researchers and trainers to systematically assess the efficacy of interfaces and work processes in support of enhanced individual and team performance. Such an approach to the development and implementation of a configurable STE is predicated on the belief that distributed task environments should be subject to systematic and objective processes analogous to those used in systems engineering.

The proposed methodology to assess team performance represents a departure from current research norms. It relies upon the use of time windows in an STE to populate points within a task complexity space. This comprehensive architecture will allow one to evaluate many different dimensions upon which team performance can vary (e.g., team roles, communication methods). This architecture incorporates the objective team performance measures discussed previously (e.g., evaluation of time-critical team interactions–team time windows). Such an architecture will not only allow researchers to understand team interaction at any one point in the task space, but more importantly allow one to investigate the changes in team interaction that corresponds to moving away from the origin of the complexity space.

## References

Bondanella, J.R., Lewis, M., Steinberg, P.S., Park, G.S., Levy, D.G., Ettedgui, E., Oaks, D.M., Sollinger, J.M., Winkler, J.D., Halliday, J.M. and Way-Smith, S., 1999, *Microworld Simulations for Command and Control Training of Theater Logistics and Support Staffs* (RAND: Washington, D.C.).

Booch, G., Rumbaugh, J. and Jacobson, I., 1999, *The Unified Modeling Language User Guide* (Reading, MA: Addison-Wesley).

Brannick, M.T., Salas, E. and Prince, C., 1997, *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (Mahwah, NJ: Lawrence Erlbaum Assoc.).

Bul, T.X. and Slvasankaran, T.R., 1990, Relation between GDSS use and group task complexity. In *Proceedings of the 23rd Annual Hawaii International Conference on System Sciences*, J.F. Nunamaker (Ed.), Jr. 23, pp. 69–78 (Los Alquitos, CA: IEEE Computer Society Press).

Byström, K. and Järvelin, K., 1995, Task complexity affects information seeking and use. *Information Processing and Management*, **31**, 191–213.

Campbell, D.J., 1991, Goal levels, complex tasks, and strategy development: a review and analysis. *Human Performance*, **4**, 1–31.

Campbell, D.J., 1988, Task complexity: a review and analysis. *Academy of Management Review*, **13**, 40–52.

Cannon-Bowers, J.A. and Salas, E., 1998, *Making Decisions Under Stress: Implications for Individual and Team Training* (American Psychological Association: Washington, D.C.).

Cannon-Bowers, J.A. and Salas, E., 1997a, Teamwork competencies: the intersection of team member knowledge, skills, and attitudes. In *Workforce Readiness*: *Competencies and Assessment*, H.F. O'Neil (Ed.), pp. 151–174 (Erlbaum: Hillsdale, NJ).

Cannon-Bowers, J.A. and Salas, E., 1997b, A framework for developing team performance measures in training. In *Team Performance Assessment and Measurement: Theory, Methods, and Applications*, M. Brannick, E. Salas and C. Prince (Eds.), pp. 45–62 (Hillsdale, NJ: Erlbaum).

Chiles, J.R., 2001, *Inviting Disaster: Lessons from the Edge of Technology* (New York: HarperBusiness).

Daft, R.L. and Macintosh, N.B., 1981, A tentative exploration into the amount and equivocality of information processing in organizational work units. *Administrative Science Quarterly*, **26**, 207–224.

Dennis, A.R., Tyran, C.K., Vogel, D.R. and Nunamaker, J.F., Jr., 1990, An evaluation of information technology to support strategic management. *Proceedings of the 11th International Conference on Information Systems*, **11**, 35–52.

Doyle, J., 1979, Truth Maintenance System. *Artificial Intelligence*, **12**, 231–271.

Fleishman, E.A. and Zaccaro, S.J., 1992, Toward a taxonomy of team performance functions. In *Teams: Their Training and Performance*, R.W. Swezey and E. Salas (Eds.), pp. 31–56 (Norwood, NJ: Ablex).

Furuta, K. and Kondo, S., 1993, An approach to assessment of plant man-machine systems by computer simulation of an operator's cognitive behavior. *International Journal of Man-Machine Studies*, **39**, 473–493.

Gallupe, R., DeSanctis, G. and Disckson, G., 1988, Computer-based support for group problem-finding: an experimental investigation, *MIS Quarterly*, 277–295.

Gray, W.D., 2002, Simulated task environments: the role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, **2**, 205–227.

Handley, H.A.H., Zaidi, Z.R. and Levis, A.H., 1999, The use of simulation models in model driven experimentation. *Systems Engineering*, **2**, 108–128.

Hammond, K.R., 1986, Generalization in operational contexts: what does it mean? Can it be done? *SMC*, **16**, 428–433.

Hammond, J.M., 1998, *Distributed Collaborative Design Teams: Media Effects on Design Processes*, MS thesis, Purdue University.

Harvey, C.M., 1997, *Toward a Model of Distributed Engineering Collaboration*, PhD thesis, Purdue University.

Harvey, C. M., 2001, Gauging team tasks: can one improve the process? In *Proceedings of the Summer Computer Simulation Conference 2001* (Orlando, FL), pp. 391–396.

Hiltz, S.R., Johnson, K. and Turoff, M., 1987, Experiments in group decision making: communication process and outcome in face-to-face versus computerized conferences, *Human Communication Research*, **13**, 225–252.

Howie, D.E. and Vicente, K.J., 1998, Measures of operator performance in complex, dynamic microworlds: advancing the state of the art. *Ergonomics*, **41**, 485–500.

Jasek, C.A. and Jones, P.M., 2001, Cooperative support for distributed supervisory control. In *Coordiantion Theory and Collaboration Technology*, G.M. Olson, T.W. Malone and J.B. Smith (Eds.), pp. 311–339 (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.).

Laudeman, I.V. and Palmer, E.A., 1995, The quantitative measurement of observed workload in the analysis of aircrew performance, *The International Journal of Aviation Psychology*, **5**, 187–197.

Lewis, F., 1982, *Facilitator: A Microcomputer Decision Support System for Small Groups*, PhD thesis, University of Louisville.

Pew, R.W. and Mavor, A.S., 1998, *Modeling Human and Organizational Behavior: Application to Military Simulations* (Washington, D.C.: National Academy Press).

Raby, M. and Wickens, C.D., 1994, Strategic workload management and decision biases in aviation, *The International Journal of Aviation Psychology*, **4**, 211–240.

Rothrock, L., 2001, Using time windows to evaluate operator performance. *International Journal of Cognitive Ergonomics*, **5**, 95–119.

Sanderson, P.M., 1993, *Designing for Simplicity of Inference in Observational Studies of Process Control: ESDA and MacSHAPA*. Paper presented at the *Fourth European Meeting on Cognitive Science Approaches to Process Control* (CSAPC'93), Hillerod, Denmark.

Sauer, J., Wastell, D.G. and Hockey, G.R.J., 2000, A conceptual framework for designing microworlds for complex work domains: a case study of the cabin air management system. *Computers in Human Behavior*, **16**, 45–58.

Smith-Jentsch, K.A., Johnston, J.H. and Payne, S.C., 1998, Measuring team-related expertise in complex environments. In *Making Decisions Under Stress: Implications for Individual and Team Training*, J.A. Cannon-Bowers and E. Salas (Eds.), pp. 61–87 (Washington, D.C.: American Psychological Association).

Watson, R., DeSanctis, G. and Poole, M., 1988, Using a GDSS to facilitate group consensus: some intended and unintended consequences, *MIS Quarterly*, **12**, 463–477.

Weeks, G.D. and Chapanis, A., 1976, Cooperative versus conflictive problem solving in three telecommunication modes, *Perceptual and Motor Skills*, **42**, 879–917.
Williams, E., 1977, Experimental comparison of face-to-face and mediated communication: a review, *Psychological Bulletin*, **84**, 963–976.
Wood, R.E., 1986, Task complexity: definition of the construct, *Organizational Behavior and Human Decision Processes*, **37**, 60–82.

## About the authors

**Ling Rothrock** is an assistant professor at the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at the Pennsylvania State University. He earned his PhD in industrial engineering from Georgia Institute of Technology in 1995 and served as an officer in the United States Army until 2000. He then began his academic career as a faculty member at Wright State University before joining the Penn State faculty in 2002. Dr Rothrock's research areas include human performance assessment and modeling as well as human-in-the-loop simulations.

**Craig Harvey** is an assistant professor in the Department of Industrial and Manufacturing Systems Engineering at Louisiana State University (LSU). Dr Harvey served as an officer in the United States Air Force until 1992. After leaving the Air Force, he worked for Sallie Mae and KnowledgeWare. Upon leaving industry in 1994, he earned his PhD in industrial engineering from Purdue University in 1997. Dr Harvey has held academic positions at the University of Oklahoma and Wright State University before joining the faculty at LSU in 2002. Dr Harvey's research areas include team collaboration in complex settings, usability engineering, and human-computer interaction.

**John Burns** is a Vice President and Principal Analyst at Sonalysts, Inc. His research interests include human performance measurement, team training, and advanced training system development. The primary focus of Burns' research is to take advantage of recent advancements in team and individual training methodologies and the emergence of cutting-edge technologies to support instruction and training at the time of need and in the most effective manner possible. Dr Burns received his PhD in psychology from the University of Massachusetts in Amherst and is an active participant in the Human Factors and Ergonomics Society and the Department of Defense Human Factors Technical Advisory Group.