A Framework for Developing and Using Shared Mental Models in Human-Agent Teams

Matthias Scheutz, Tufts University, **Scott A. DeLoach**, Kansas State University, and **Julie A. Adams**, Vanderbilt University

Converging evidence from psychology, human factors, management and organizational science, and other related fields suggests that humans working in teams employ shared mental models to represent and use pertinent information about the task, the equipment, the team members, and their roles. In particular, shared mental models are used to interact efficiently with other team members and to track progress in terms of goals, subgoals, achieved and planned states, as well as other team-related factors. Although much of the literature on shared mental models has focused on quantifying the success of teams that can use them effectively, there is little work on the types of data structures and processes that operate on them, which are required to operationalize shared mental models. This paper proposes the first comprehensive formal and computational framework based on results from human teams that can be used to implement shared mental models for artificial virtual and robotic agents. The formal portion of the framework specifies the necessary data structures and representations, whereas the computational framework specifies the necessary computational processes and their interactions to build, update, and maintain shared mental models.

Keywords: shared mental models, formal computational framework

INTRODUCTION

As societies increasingly embrace pervasive immersive technologies (e.g., smartphones and activity/health monitors), new forms of collaboration are emerging that require humans to quickly form teams to tackle problems and achieve common goals. These teams can range from local ad hoc groups formed to solve a currently pressing problem (e.g., bystanders rescuing passengers from burning cars after a multicar crash) to long-standing well-trained professionals that provide specialized services (e.g., first responders searching for survivors in collapsed buildings). Although coordinating team efforts may be largely face-to-face in the first case, the use of phones and other communication devices is critical for the latter. The teaming structure in the former is flat and event driven but, in the latter, based on a well-established hierarchy and prior training. Common to both scenarios—and any type of teaming effort in general—are two questions: What is required to be an effective team? and How can technology support teaming and improve team performance?

Prior behavioral research has demonstrated that human teams coordinate their activities more effectively and achieve better overall task performance when team members track and take into account one another's goals, intentions, beliefs, as well as other performance and task-related states—that is, when they use a "shared mental model" (SMM; Cannon-Bowers, Salas, & Converse, 1993; Cooke et al., 2003; M. Lee, Johnson, & Jin, 2012; Mathieu, Goodwin, Heffner, Salas, & Cannon-Bowers, 2000; Mohammed, Ferzandi, & Hamilton, 2010). SMMs are related to and can be subsumed under the concept of "mental model," which typically refers to the types of hypothesized knowledge structures that humans build to make sense of their world, to make inferences based on the available information, and to make predictions about future states (e.g., Held, Knaff, & Vosgerau, 2006; Johnson-Laird, 1983; Rouse & Morris, 1986). Although mental model research in psychology has focused on using mental models to explain various types of human reasoning (Borgman, 1986; Gray, 1990), mental models in the

Address correspondence to Julie A. Adams, Vanderbilt University, 2301 Vanderbilt Place, Station B #351824, Nashville, TN 37235, USA, julie.a.adams@vanderbilt.edu.

Journal of Cognitive Engineering and Decision Making
201X, Volume XX, Number X, Month 2017, pp. 1–22
DOI: 10.1177/1555343416682891

context of teams have more to do with establishing and maintaining common ground (e.g., Clark, 1996) and building team mental models (e.g., Cannon-Bowers et al., 1993) that aid in decision making and the adjustment of one's behavior based on predictions about the other team members' future activities and actions. Thus, team mental models are critical for making sense of team activities, for understanding the dynamic changes of team goals and team needs (including the possibly dynamic changes of roles and functions of team members), and for tracking the overall moment-to-moment state of the team.

Given the importance of SMMs for human teams, it is natural to hypothesize that similar SMM capabilities in artificial agents (e.g., virtual or robotic agents) may improve the efficiency and productivity of mixed human-agent teams—we call this the *shared mental model hypothesis for artificial agents*. Substantiating and evaluating the hypothesis requires the implementation of SMMs in artificial agents and ultimately performing experiments with mixed human-agent systems conducting typical team tasks. However, no current control architecture for artificial agents is capable of using SMMs, and Wizard of Oz–style evaluations (where artificial agents are teleoperated by humans) can be conducted in only a very limited fashion for realistic tasks. Thus, making progress toward developing (and eventually evaluating) SMM capabilities in artificial agents requires a better understanding of the representations and processes underwriting SMMs in humans. Furthermore, how those representations and processes can be captured as computational structures must be investigated—that is, what detailed information SMMs contain about team members, including their intentions, goals, knowledge, beliefs, capabilities, activities, and performance factors. The aim of this paper is to provide the formal and computational foundations for integrating SMMs into agent architectures for artificial agents to be used in mixed human-agent teams.

We begin by reviewing the SMM literature and relevant features of humans' SMMs. A typical mixed human-agent scenario, in which SMMs can improve team coordination and performance, is described to directly compare artificial agents with and without SMMs. A general formal SMM framework in introduced that lays out the necessary SMM concepts and rules for those concepts to allow an agent to capture and reason with information contained in SMMs. We describe how this formal framework can be realized computationally via a computational architecture that has various functional components necessary to implement and use SMMs in artificial agents. Finally, representative partial implementations of the formal framework and its application to the human-agent scenario are provided.

A BRIEF OVERVIEW OF SMMs

The term *mental model* refers to hypothesized knowledge structures that humans use to reason about the world, to make inferences based on the available information, and to make predictions about future states (e.g., Held et al., 2006; Johnson-Laird, 1983; Rouse & Morris, 1986). Proposals exist regarding how mental models are structured (e.g., Held et al., 2006; Johnson-Laird, 1983), but much of the research focuses on understanding human reasoning based on mental models (e.g., Borgman, 1986; Gray, 1990). The general concept of mental models was extended to teams based on the notion that team members have elements of their individual mental models in common: the SMMs (e.g., Cannon-Bowers et al., 1993; Cooke et al., 2003). SMMs aid decision making and adjustment of one's behavior based on predictions of other team members' current state, future activities, and actions. Thus, SMMs are critical for making sense of team activities and understanding the dynamic changes in team goals and needs, team members' roles and functions, and the overall team state.

The early SMM research reflected the belief that individuals' SMMs were composed of multiple models (e.g., Cannon-Bowers et al., 1993), which incorporated the equipment model, the task model, the team interaction model, and the team model. The equipment model represents the equipment used for the task, whereas the task model represents the task. Two models are used to represent the team: the team interaction model incorporates the team members' roles, responsibilities, and

TABLE 1: The Four Components of a Shared Mental Model

Task Model		Team Model	
Equipment	Task	Team Interaction	Team (Teammates')
Equipment functioning	Task procedures	Roles/responsibilities	Knowledge
Operating procedures	Likely contingencies	Information sources	Skills
Likely failures	Environmental constraints	Interaction patterns	Performance history
Equipment/system limitations	Task strategies	Information flow	Tendencies
	Likely scenarios	Communication channels	Preferences
	Task component relationships	Role interdependencies	Attitudes

interactions, while the team model represents characteristics unique to each team member, such as knowledge, skills, and abilities.

The more recent literature consolidates the four early SMM components into only the task and team models (e.g., M. Lee et al., 2012; Mathieu et al., 2000; Mathieu, Rapp, Maynard, & Mangos, 2009; Mohammed et al., 2010). The task model represents the task and how the environment can affect the task demands. The team model incorporates the representation of the team members and their interaction with one another. Mathieu et al. (2000) proposed that the team SMM incorporates the originally proposed team interaction and the team models, whereas the task SMM incorporates the originally proposed task and equipment models. M. Lee, Johnson, Lee, O'Connor, and Khalil (2004) proposed that a SMM has five components: team knowledge, team skills, team attitudes, team dynamics, and team environment. However, more recently, they too have adopted the team and task SMM representation (M. Lee et al., 2012).

Table 1 provides an overview of the categorizations. Although the task components are directly related to the specific task, they do not change frequently. The associated equipment components rarely change; thus, the task model is relatively stable. The team interaction components change depending on the tasks but are moderately stable. However, the team components are the most dynamic due to the direct relationship to individual team members' performance factors and the task situation. As a result, the team model has low stability.

The primary research focus has been on the behavioral phenomenon of SMMs, the mental model's impact on team performance, and associated metrics (e.g., Banks & Millward, 2000; Cooke et al., 2003; DeChurch & Mesmer-Magnus, 2010; Espevik, Johnsen, & Eid, 2011; Lim & Klein, 2006). Some have suggested that SMMs are composed of schemas (e.g., DuRussell & Derry, 2005), whereas others suggest alternative representations (e.g., M. Lee et al., 2012; Wilson, Salas, Priest, & Andrews, 2007).

Even though the behavioral literature largely omits SMM representations that translate into computational data structures and algorithms for maintaining and accessing SMMs, several principles for computational SMM models are nevertheless implied according to how humans realize SMMs:

Consistency: The primary goal is maintaining consistency by resolving conflicts due to (a) differing perceptions, (b) differing knowledge states, (c) asynchronous information, and (d) missing updates.

Reactivity: Effective teams quickly react to unanticipated events or state changes by informing team members of the changes and adapting goals and plans to account for the new situations.

Proactivity: Effective teams anticipate problems, bottlenecks, and failures and take proactive actions, such as asking for clarification or offering assistance.

Coordination: Effective teams excel at coordinating actions via overall cooperative attitudes, such as establishing joint goals and plans, transparent task assignments, and truthful information sharing.

Knowledge stability: Humans understand the stability of information (e.g., team members' location) over time and adjust their sampling rates according to their confidence in the information's validity.

Although artificial intelligence, robotics, and human-robot interaction researchers have worked toward designing mixed human-robot teams incorporating some of these principles—belief-desires-intentions or theory of mind representations (e.g., Bosse, Memon, & Treur, 2007; Castelfranchi, 1998; Fong, Nourbakhsh, & Dautenhahn, 2003; Georgeff, Pell, Pollack, Tambe, & Wooldridge, 2003) as well as coordination rules and mechanisms and task assignment algorithms (e.g., Dahl, Matarić, & Sukhatme, 2009; Luo, Chakraborty, & Sycara, 2015; Tambe, 1997)—there is currently no comprehensive computational SMM addressing them all, let alone an implemented robotic system.

The software agent literature, for example, provides a conceptual SMM ontology (Jonker, van Riemsdijk, & Vermeulen, 2010), applies the concept of team member roles (Zhang, 2008), and proposes a game-theoretic SMM (Richards, 2001); however, no fundamental details are provided. An SMM that couples a software agent with a human completing a simplistic computer-based task predicted the human's cognitive load (Fan & Yen, 2007, 2011). This SMM has limited scalability to larger teams distributed across a broader set of teaming structures and domains due to the fundamental assumption that every human interacts with an associated agent via a graphical user interface.

Some of the underlying concepts related to understanding how to represent the human SMM construct as a computation framework were represented as an SMM ontology, which the authors clearly state is not intended "as a design for implementation" (Jonker et al., 2010). Their intent was to gain an understanding of the essential SMM concepts and the relationships among the concepts; as such, our efforts share the same intent but actually define logical relationships and a software system design to support the computational SMM. Jonker and colleagues provide an example scenario with some logical relationships (Jonker et al., 2010; Jonker, Riemsdijk, & Vermeulen, 2011), but the sample problems are fairly simplistic; as a consequence, the system will likely suffer from computational limitations as the domain complexity increases. Furthermore, the representation does not appear to efficiently account for uncertainty.

The human-robot interaction research has focused on understanding humans' mental models of robots (e.g., Keisler & Goetz, 2002; S.-L. Lee, Keisler, Lau, & Chiu, 2005; Syrdal, Dautenhahn, Koay, Walters, & Otero, 2010; Stubbs, Wettergreen, & Hinds, 2007). Some claim to develop mental models (e.g., Miwa, Okuchi, Itoh, Takanobu, & Takanishi, 2003; Park, Kwon, & Park, 2007) or SMMs (e.g., Burkhard, Bach, Berger, Brunswieck, & Gollin, 2002) for cognitive robotics; however, these approaches do not align with the traditional human-based mental model and SMM literature. For example, Miwa et al. (2003) developed a model of mood and affect within the robot, whereas Park et al. (2007) developed a petri net containing eight characteristics of a superego that they incorrectly call a *mental model*. Similarly, it appears that Burkhard and colleagues' (2002) "mental model" is actually a representation of a world model, intentions, and planning.

Existing robotic mental models incorporate underlying assumptions that do not support domain independence or generalizability or limit the model scalability. For example, Goodrich and Yi (2013) present an SMM framework for a human-robot team where the robot completes the tasks that robots are best at—as a direct assistant, a "wingman," to the human. This approach differs significantly from the presented approach in that the framework assumes that robots must hold more complex objectives that are not necessarily in support of the human's primary tasks but may be independent task and goal responsibilities. Nikolaidis and Shah (2013) share some underlying concepts with the proposed framework and have obtained impressive outcomes with actual robot manipulators. Although the concept of cross-training to develop the SMM is a good fit for their domain, relying on cross-training is difficult for uncertain and dynamic environments that are not well understood a priori and for larger teams composed of multiple humans and heterogenous robots. The framework presented in this paper has been designed to accommodate more complex domains and teaming situations.

Others suggest developing (Neerincx, de Greef, Smets, & Sam, 2011; Ososky et al., 2012) or have developed simulated robotic SMMs (Kennedy & Trafton, 2007) but do not provide comprehensive frameworks for realizing robot SMMs. It is suggested that SMMs are necessary to make the robots true teammates (Ososky et al., 2012), and ACT-R (adaptive control of thought–rational) was used to demonstrate, in simulation, the use of mental models to improve performance (Libiere, Jentsch, & Ososky, 2013), although no comprehensive framework was provided for integrating SMMs into robots that can function in uncertain and dynamic environments.

A PROTOTYPICAL SMM APPLICATION

A mass casualty emergency provides an example for understanding the functional role of SMMs in team tasks, including what aspects are represented, what updates can be made, and, most important, how team interactions differ depending on whether SMMs are used or not. A basic assumption throughout this paper is that human-robot teams train with one another, just as human teams teams train, such as first responders. As a result, the human responders develop mental models of the robots and SMMs of their team. Furthermore, such training allows the robots to develop SMMs for use during response deployments.

Imagine that 50,000 people attending a National Football League game are exposed to a biological contaminant that can cause death after 48 hours (Humphrey & Adams, 2009, 2011, 2013). Approximately 75% of the victims require immediate care, whereas the remaining victims can wait for care (delayed care). Victims requiring immediate care receive some treatment in the contaminated area (hot zone) and are transported to an off-site care facility. The temperature is in the low 80s (Fahrenheit), and the humidity level is 80%.

The direct human teammates (*H*; Humphrey, 2009; Scholtz, 2003) reside in the hot zone and interact with the ground robots (*R*) to achieve the assigned tasks and ensure effective team interaction. The Level B personal protective equipment—including a breathing apparatus, hooded chemical-resistant clothing, gloves, and boots—limits the humans' field of view and causes increased body temperature. Furthermore, the breathing apparatus limits the deployment to approximately 2.5 hr.

Assume that the paramedic team includes (a) two robots that can transport supplies and triage victims and (b) two direct human teammates (H_1 and H_1). Furthermore, assume robot R_1 is triaging victim *V*, who requires immediate care that R_1 cannot provide. R_1 requests assistance from the closest human, H_1, who is triaging other victims. R_2 and H_2 are working nearby.

The following scenarios demonstrate robots with and without SMMs. As compared with robots without SMMs, robots possessing SMMs can perceive changes in human responders' performance and the current situation, and they can be proactive in taking actions that minimize the demands and impact on the human responders while improving the overall team's performance. The scenarios highlight where the SMMs achieve these objectives (basic mental model representational primitives and human performance functions are shown in bold and underlined font, respectively; SMM capabilities shown in italic).

Scenario Without SMM	Scenario with SMM
R_1 does not know how long H_1 will need to triage V. As H_1 begins walking toward V, H_1 notes the high air temperature and humidity levels, coupled with the deployment duration and required personal protective gear, and seeks to minimize his or her physical workload by not carrying any supplies. H_1 asks R_1 whether R_1 has the necessary supplies, which it does not. H_1 instructs R_1 to find a robot nearby with the necessary supplies. R_1 begins contacting all robots in the order of their last known locations. Eventually, R_1 contacts R_2 to request supplies, and R_2, in turn, contacts H_2, with whom R_2 is working, for permission to provide R_2's supplies. H_2 determines that there are sufficient supplies and gives R_2 permission; R_2 moves toward R_1's position. H_1 is already waiting at that location to retrieve the supplies. H_1 determines that it will take approximately 10 min to treat V. R_1 waits for H_1 to provide a plan for R_1 to proceed with triaging additional victims.	R_1 **believes** that H_1 is **capable** of triaging V, and based on prior training and task predictions, R_1 **knows** it takes 10 min to treat V. R_1 **perceives** the high air temperature and humidity levels, coupled with the deployment duration and required personal protective gear, and instantiates a **goal** to minimize H_1's physical workload by not requiring H_1 to carry supplies to V's position. R_1 *does not have the necessary supplies, but* **knows** *that* R_2 *has them. Since* R_2 **knows** *that* H_2 *has sufficient supplies, the robots determine proactively that* R_2 *will* **adopt a goal** *to bring the supplies to* R_1. Managing H_2's workload, R_2 informs H_2 of the plan before delivering the supplies, while R_1 **requests** H_1's assistance and proactively **communicates** that it will have the necessary supplies. R_1 also manages H_1's cognitive workload by developing a **plan** to triage additional victims while H_1 triages V, and R_1 **communicates** the plan, which H_1 acknowledges and R_1 **adopts**.

Specifically, the scenarios demonstrate that R_1 without a SMM requires more communication and increases H_1's and H_2's workloads while decreasing overall team task performance due to the lack of (proactive) concurrent activities. Conversely, robots with SMMs minimize H_1's and H_2's distractions, interruptions, and workloads via proactive communication and planning while increasing the team's overall performance (italicized text). Thus, the scenario demonstrates the difference in performance enabled by SMMs while highlighting the necessary representations and processes needed in agents to build, maintain, and utilize SMMs. The next two sections introduce the representational and computational frameworks required for general SMM processing in artificial agents.

A GENERAL FORMAL FRAMEWORK FOR SMMS

From a computational perspective, SMMs consist of two key elements: first, data representations that capture state information about the team, tasks, and environment, as well as task-relevant knowledge that is shared among team members, such as facts, rules, procedures, principles, and obligations; second, computational processes that create and maintain the data representations. Although the former has components that vary from task to task (e.g., the equipment used, the task goals, the team composition), the latter provides general mechanisms that are applicable across tasks (e.g., how to update the belief of another team member based on observations or communicated information). Developing algorithms for using SMMs in artificial agents requires a comprehensive formal framework that provides intuitive mappings of formal constructs to existing human SMM structures and extends the elements traditionally included in SMMs by novel structures that account for human performance functions. We significantly extend our previous pragmatic and mental modeling framework (Briggs & Scheutz, 2011, 2012) to comprehensively capture all relevant aspects of SMMs reported in the literature, as well as novel components based on human performance functions.

The following variables are used throughout the formal framework component specification:

ϕ_i used for formulas, α_i for actions and skills, π_i for plans, γ_i for goals and subgoals, σ_i for situations, τ_1 for object types, η_1 for events, and A_1 for agents (which refer to human and artificial agents alike). Capturing uncertain knowledge is straightforward, and some examples are presented in the following sections, but a thorough presentation is beyond the scope of this manuscript.

Basic Representational Primitives

The formal framework encompasses five comprehensive sets of predicates that capture different aspects of agents, tasks, and environments relevant to SMMs:

- *F1:* Agent capabilities and propensities—including perceptions, actions, skills, traits, rank, and possibly other relevant agent properties
- *F2:* Moment-to-moment agent and task states—including knowledge, belief, and human performance states; adopted goals and plans; and ongoing activities
- *F3:* Known and accepted obligations and norms pertaining to the task and performance domains (including rules for goal and activity negotiations) and general norms about agent behavior
- *F4:* Activity and equipment types
- *F5:* Functional roles of agents in teams, in terms of the activities that they ought to perform

F1: Agent capabilities and propensities. Agent capabilities and propensities require defining different aspects of the general notion CAPABLE(A,X), which states that agent A is capable of carrying out X, where X can be an action, a skill, or a plan. This general notion of capability can be contrasted with a situation-specific version stated in terms of goal states: CAPABLE(A,ϕ,σ), which means that agent A can bring about the state ϕ in situation σ. For example, the agent may be capable of triaging a victim: CAPABLE(A, *triage–victim*).

Analogous to capabilities for behavior, PERCEIVABLE(A,X) means that agent A can perceive X, where X can be a type of agent, object, event, or activity (e.g., a skill or plan being carried out by another agent). Again, analogous to capabilities, a notion extended by situations is introduced, PERCEIVABLE(A,ϕ,σ), which means that agent A can perceive whether ϕ is true in situation σ. Capturing imperfections in perception can use PERCEIVABLE$(A,\phi,\sigma)_p$, where p is the probability of agent A perceiving whether ϕ is true in situation σ.

We also represent general agent propensities for different situations: TENDSTO(A,X,σ), which means that agent A has a tendency to perform X in situation σ. For example, a generally polite agent will say “hello” when seeing another agent:

$$\text{TENDSTO}(A_1, \text{SAY}(A_1, A_2, \text{‘hello’}), \{\text{SEES}(A_1, A_2)\}).$$

The definitions can be extended to capture an agent’s likelihood to exhibit a particular propensity, by adding probabilities as an argument. Probability distributions over agent behaviors in different situations can be obtained from these augmented definitions to predict what agents are likely to do.

F2: Agent and task states. Agent states require developing a sufficiently comprehensive account of what humans track about their teammates. Therefore, predicates for cognitive states (e.g., beliefs and goals) and noncognitive states (e.g., workload) provide a sufficiently comprehensive account of what humans track in relation to their teammates.

Knowledge and beliefs: In addition to the basic distinction between “knowledge” and “belief” states (i.e., KNOWS(A,ϕ) versus BELIEVES(A,ϕ)) and their generalizations (COMMON KNOWLEDGE(ϕ) and COMMONBELIEF(ϕ)), we use KNOWS-OF(A,X), where X can be any type of entity—agent, object, event, activity, plan, goal, and so on; “intends to know”, ITK(A,X), where X can be any type of agent, object, event, activity, plan, goal, and so on; and KNOWS-HOW(A,X), meaning that agent A knows how to perform X, where X is an action, a skill, or a plan.

Goals and plans: Predicates are necessary for the achievement and maintenance of goals, GOAL(A,γ); subgoals, SUBGOAL(γ_i,γ_j); and common goal, CG(γ). Plans are sequences of actions and skills that achieve a goal in a given situation: ACHIEVES(π,ϕ,σ). Agents can adopt plans, ADOPTED(A,π,σ), and execute them, EXECUTING(A,π,σ), in given situations, σ, where adopting implies that the agent is trying its best to execute the plan in σ or some subsequent situation. We introduce the notion "seeing to it that ϕ" STI(A,ϕ)), which means that agent A is carrying out some procedures that attempts to establish ϕ (the standard logical notations are defined at http://www.rapidtables.com/math/symbols/Logic_Symbols.htm):

$$\mathrm{STI}(A,\phi,\sigma) :\leftrightarrow \exists\pi[\mathrm{ADOPTED}(A,\pi,\sigma) \\ \wedge \mathrm{ACHIEVES}(\pi,\phi,\sigma)).$$

Finally, we introduce the effect predicate EFFECT(π,ϕ,σ), which means that executing π establishes ϕ in σ. Probabilities can be added if stochastic domains are to be considered—for example, $\mathrm{EFFECT}(\pi,\phi,\sigma)_p$ where p is the probability that ϕ will be achieved when π is executed in σ.

Affective states: Human performance moderator functions can significantly affect overall team interactions and performance; thus, we allow for modeling noncognitive states, such as affect and visceral states. Specifically, qualitative indications are allowed for an agent's level of exhaustion EXHAUSTION(A,L), where L is the fatigue level of agent A, and workload WORKLOAD(A,L). Additional noncognitive states can be added as needed. The important point is to be able to include states that can be measured during task performance and whose measurements may be available to team members, which in turn allows them to draw inferences about other team members' performance. For example, if an agent's workload is high, then no additional goals are to be given to that agent, and if an agent's level of exhaustion is high, it may no longer be able to carry out all activities of which it is capable.

F3: Norms and obligations. Norms and obligations are defined in terms of (modal) operators O, "obligatory," and P, "permissible," to capture rules regarding negotiation of goals and assignments, obligatory normative behaviors (e.g., a first responder triaging a victim), and other task-based constraints. For example, PROPOSES(A_1,A_2,X) can indicate that agent A_1 proposes X to agent A_2, where X can either be a plan or a goal, thus effectively asking whether A_2 will adopt X. In response, A_2 either REJECTS (A_2,A_1,X)) or ACCEPTS(A_2,A_1,X)); in the latter case, A_2 has committed to carrying out X, if X is a plan (i.e., ADOPTED(A_2,X,σ)) or A_2 has a GOAL (A,X). We also introduce the predicate SUPERIOR(A_1,A_2) to indicate that A_1 is higher up in the command hierarchy than A_2.

F4: Activity and equipment types. Activity and equipment types are specified in terms of their pre- and postconditions. For example, a "victim retrieval" activity can be defined as being capable of carrying victims out of danger zones (where σ′ is a situation after σ):

$$\mathrm{ACTIVITY}(\mathit{victim\_retrieval}) \colon \leftrightarrow \forall A,V, \\ \sigma[\mathrm{AGEMT}(A) \wedge \mathrm{VICTIM}(V) \wedge \mathrm{CAPABLE}(A,\mathit{carry},V) \wedge \\ \mathrm{INDANGERZONE}(V,\sigma) \wedge \\ \mathrm{PERFORMSON}(A,\mathit{carry},V,\sigma) \rightarrow \mathrm{SAFE}(V,\sigma')].$$

Equipment types are defined in terms of the actions that can be performed with the equipment. For example, a stretcher is an object that can be used (by a capable agent) for victim retrieval:

$$\mathrm{STRETCHER}(x) :\leftrightarrow \exists y,A[\mathrm{ACTION}(y) \wedge \\ \mathrm{AGENT}(A) \wedge \mathrm{CAPABLE}(A,y) \wedge \forall\sigma,V(\mathrm{VICTIM}(V) \wedge \\ \mathrm{PERFORMSONUSING}(A,y,V,x,\sigma) \\ \rightarrow \mathrm{RETRIEVED}(V,\sigma'))].$$

The specific actions to be performed with a stretcher are not defined, since many actions can be performed with it to retrieve a victim (e.g., putting the victim on it and dragging it to a safe location). Generic equipment needs for goals and equipment required by agents are expressed via REQUIRED(e,ϕ) (i.e., the equipment e is required to achieve goal ϕ) and REQUIRES(A,e) (i.e., agent A requires equipment e).

F5: Functional roles of agents in teams. Functional roles of agents in teams are defined in terms of goals, equipment requirements, necessary capabilities, obligations, permissible and impermissible states, and actions of the agent assuming the role. For example, the role of a searcher can be defined as follows:

$$
\begin{aligned}
&\mathrm{HASROLE}(x, searcher) :\leftrightarrow \\
&\quad \mathrm{GOAL}(x, searchArea_7) \wedge \\
&\quad \mathrm{GOAL}(x, reportAllResults) \wedge \ldots[\text{add all goals}] \\
&\quad \mathrm{REQUIRES}(\mathrm{x,map}) \ \wedge \ldots[\text{additional requirements}] \\
&\quad \exists y \mathrm{CAPABLE}(x, y) \wedge \\
&\quad \mathrm{ACTIONTYPE}(x, door-opening) \\
&\qquad \wedge \ldots[\text{add. actions}] \\
&\quad \forall z \mathrm{SUPERIOR}(z, x) \rightarrow \\
&\quad \mathrm{OBLIGATED}(x, informed(z)) \\
&\qquad \wedge \ldots[\text{add. obligations}] \\
&\quad \mathrm{PERMISSIBLE}(a, \phi_1) \wedge \ldots \begin{bmatrix} \text{additional permissible} \\ \text{states and actions} \end{bmatrix} \\
&\quad \neg \mathrm{PERMISSIBLE}(a, \psi_1) \wedge \ldots \begin{bmatrix} \text{additional impermissible} \\ \text{states and actions} \end{bmatrix}.
\end{aligned}
$$

A hierarchically structured team can be defined in terms of an agent's roles, the command structure, and the available equipment types:

$$
\begin{aligned}
&\mathrm{TEAM}(x, rescue-team) :\leftrightarrow \\
&\quad !\exists x \exists y, z \mathrm{HASROLE}(x, searcher) \wedge \\
&\quad \mathrm{MEDIC}(y) \wedge \ldots[\text{additional roles}] \\
&\quad \mathrm{SUPERIOR}(z, x) \wedge \mathrm{SUPERIOR}(z, y) \\
&\quad \wedge \ldots[\text{additional command relationships}] \\
&\quad \exists k \mathrm{MEDICALKIT}(k) \wedge !\exists m \mathrm{MAP}(m) \\
&\quad \wedge \ldots[\text{additional equipment}].
\end{aligned}
$$

Agent roles and equipment types are filled by the specific agents assigned to those roles and the available equipment.

General Rules and Update Principles

Overall, updates to an agent's SMM can be triggered by various events and mediated through the agent's perceptual system and internal state changes. For example, agent A_1 perceives a new task-relevant object, which triggers the instantiation of a new goal. Agent A_2 observing agent A_1 needs to update its model to accommodate agent A_1's new goal. One can define general principles for perceptions and actions that allow agents to make quick inferences and updates to their SMM based on their perceptions of their own and other agents' actions and communications. For example, in general, all team members are truthful with each other; that is, agent A_1 believes what it hears from any team member A_2 (including itself):

$$
\begin{aligned}
&\mathrm{SAID}(A_2, \phi) \wedge \mathrm{HEARD}(A_1, \phi) \wedge \mathrm{INTEAM}(A_2, \\
&\quad \mathrm{TEAMOF}(A_1)) \rightarrow \mathrm{BELIEVES}(A_1, \phi).
\end{aligned}
$$

This principle allows an agent to update its SMM based on any communications in which it is directly involved or that it overhears. Similar principles can be defined for actions that other agents perform or for perceptions by other agents.

Additional rules can be added for interacting with other team members about adopting and dropping plans and goals, as plans and goals can change or different goal assignments may become necessary during task execution (e.g., because of unexpected events or the loss of capabilities by team members).

For example, a principle can be defined that proposed plans to achieve common goals are adopted if no previous plans exist:

$$
\begin{aligned}
&\mathrm{EFFECT}(\pi, \phi) \wedge \mathrm{CG}(\phi) \wedge \mathrm{PROPOSES}(A_1, A_2, \pi) \\
&\quad \wedge \neg \exists \pi' (\mathrm{ADOPTED}(A_2, \pi') \\
&\quad \wedge \mathrm{ACHIEVES}(\pi', \phi)) \rightarrow \mathrm{ADOPTED}(A_2, \pi).
\end{aligned}
$$

This predicate allows the formulation of more refined principles representing when agents accept goal assignments—for example, plans ordered by superiors must be adopted, as long as they do not violate common goals:

$$
\begin{aligned}
&\neg \mathrm{EFFECT}(\pi, \neg \phi) \wedge \mathrm{CG}(\phi) \wedge \mathrm{SUPERIOR}(A_1, A_2) \wedge \\
&\quad \mathrm{PROPOSES}(A_1, A_2, \pi) \rightarrow \mathrm{ADOPTED}(A_2, \pi).
\end{aligned}
$$

The above can be augmented by making it explicit that an agent has to be available to pursue a goal, AVAILABLEFOR(A,ϕ).

Principles can be defined to explicitly track the interaction between assignments and cognitive states—for example, when agent A_1 proposes a goal γ to agent A_2 and agent A_2 accepts it, then both agents believe that A_2 has that goal:

$$
\begin{aligned}
&\mathrm{PROPOSES}(A_1, A_2, \gamma) \wedge \mathrm{ACCEPTS}(A_2, A_1, \gamma) \rightarrow \\
&\quad \mathrm{BELIEVES}(A_1, \mathrm{GOAL}(A_2, \gamma)) \\
&\quad \wedge \mathrm{BELIEVES}(A_2, \mathrm{GOAL}(A_2, \gamma)).
\end{aligned}
$$

In some cases, it may be necessary to explicitly add capabilities to the principles (i.e., requiring for each planned action that the agent must be capable of performing the action before it can accept the goal). In other cases, when the agent acquires new capabilities during task execution, such explicit

TABLE 2: The Human Performance Functions Organized by Category

Acquired	Dynamic	Designed	Environmental
Training and experience Skills Prior performance on similar tasks	Physical workload Cognitive workload	*Physical workload* Organizational structure	Personal protective gear Environmental humidity Noise level Environmental temperature

Note. Italics indicate secondary categorization of a human performance function.

requirements are not necessary before accepting goals; however, the agent must notice that it does not have a particular capability to be able to request information about it (e.g., how to learn it).

General principles of what is obligatory, what is permitted, and what is not permitted are also required:

$$\text{OBLIGATED}(A,\gamma,\sigma) :\leftrightarrow \text{O STI}(A,\phi,\sigma),$$

which means that agent A is obligated to see to it, STI(A,ϕ,σ), that γ is obtained in situation σ. Similarly, predicates can be added for obligations to carry out plans, OBLIGATED(A,π,σ), and permissions to achieve goal states or carry out plans, PERMISSIBLE(A,ϕ,σ) and PERMIS SIBLE(A,π,σ), respectively. Additional modal operators for capturing temporal relationships (e.g., "in the future") can represent temporal dependencies, as in when one action is only permissible after another has been executed.

Human Performance Functions

The teaming and SMM research does not address artificial systems' prediction and assessment of human performance in real time, which can be achieved with human performance functions (HPFs). Team performance is affected by internal, environmental, task, and organizational influences (e.g., Salvucci, 2001; Weaver, Silverman, Shin, & Dubois, 2001); hence, the robots' prediction and perception of human performance in real time is key to adapting their behavior and interaction to accommodate the human.

HPFs provide one means of predicting human performance when establishing initial and updated SMMs. Derived from experimental data, HPFs determine how human performance is affected by combinations of influencing factors in differing conditions. Over 500 human performance functions have been analyzed and validated (Silverman, Johns, Cornwell, & O'Brien, 2006) for domains, such as driving (e.g., Salvucci, 2001), power plant operation (e.g., Mumaw, Roth, Vicenti, & Burns, 2000), and military applications (e.g., Weaver et al., 2001). Prior work focused on evaluating the applicability of a subset of these HPFs to human-robot interaction (e.g., Harriott & Adams, 2013; Harriott, Buford, Zhang, & Adams, 2015; Harriott, Zhang, & Adams, 2013; Harriott, Zhuang, Adams, & DeLoach, 2012).

A large number of HPFs are applicable to human-agent teams; however, only a subset is presented, as divided into four HPF categories based on the frequency of change and the human's ability to directly effect the HPF (Adams, Harriott, Zhuang, & DeLoach, 2012). Some HPFs cross categories and have a primary and a secondary categorization, as indicated in Table 2.

The acquired HPFs influence the team SMM (as defined in Table 1), are moderately stable, and are generally under the individual human's control but can suffer from individual differences. Prior performance on similar tasks represents an objective function of the agent's performance on prior, but similar tasks that can be assessed during training and real tasks. Variances in prior performance across agents may be a good predictor of future performance. Skills represent an agent's specialties acquired and developed during training and from prior experiences, whereas training and experience represent formal task training and experience or procedural knowledge obtained during training exercises or tasks. These HPFs are known a priori and can be recorded prior to completing tasks as well as after or during training.

The dynamic HPFs change frequently, are influenced by the situation, and have low stability. Cognitive workload represents the ratio of time required to do tasks to the time available (Wickens, Lee, Liu, & Gordon-Becker, 2003). Physical workload, the maximum physical work capacity, varies as a function of the duration of work and is the maximum rate of energy production during physical work (Wickens et al., 2003). These HPFs affect the task and team SMMs.

The designed HPFs represent elements that are not under the human's direct control, may be genetically inherent, and can only improve or degrade over very long periods. These HPFs can also represent human-made elements that provide overarching structures and capabilities. The organizational structure correlates with agent responsibilities, authority, supervisors, and communication channels, as in communicating only with supervisors or subordinates. The designed HPFs tend to be highly stable and affect the team and task SMMs.

The environmental HPFs are not within the team members' control, have moderate stability, and affect the team and task SMMs. Environmental HPFs may change during a mission but are generally constant or change slowly. Personal protective gear restricts human perception of and interaction with the environment. Noise level has an impact on communication among team members, which can affect understanding of task progress. Continuously high noise levels can significantly degrade human performance. Humidity and temperature can have a significant impact on human performance.

Each HPF affects the SMM and must be accounted for via predicates that capture team members' cognitive/noncognitive performance and visceral states. For example, WORKLOAD(A,L), where L is agent A' s overall workload level, represents overall workload, and OVERLOADED(A) can indicate when the overall workload level exceeds a defined task-specific threshold. Similarly, the environmental noise level affecting an agent, NOISE(A,L), and the agent's physical workload, PHYSICALWORK LOAD(A,L), can be defined. Critically, the agents will observe and track all of these states, which allow inferences about other teammates' performance. For example, if a teammate's overall workload is overloaded, the teammate will not be given a new goal, and if a teammate's physical workload level is high, the teammate can become physically fatigued and unable to complete other physical tasks. Rules and principles can be defined according to the quantitative HPF results to represent the relationships between HPFs and other parts of SMMs, such as goal assignments, plans, and so on. Although some of the principles are general in nature, others are specific to the task (e.g., how to react to high levels of workload or noise).

A COMPUTATIONAL FRAMEWORK

A scalable and extensible computational implementation of the SMM framework is required to realize the envisioned human-agent teams. This section presents the architecture and various functional components of a computational framework that applies the formal framework from the previous section to represent SMMs and to reason with them. The goal is to maintain consistent SMMs for both the humans and the agents, but consistency is easier to maintain for artificial agents.

An agent's knowledge of the current team and task states is stored in an internal SMM. As changes are made to the SMM contents, the changes must be synchronized across all agents, which includes resolving possible inconsistencies among internal SMMs from different agents. This synchronization is an instance of distributed "truth maintenance" systems (Bridgeland & Huhns, 1990), which have been widely investigated in artificial intelligence research through approaches ranging from heuristics (using the knowledge with the later time stamp; removing jointly inconsistent items and treating them as "missing knowledge") to those with formal guarantees of synchronization.

As each agent observes the environment (including other agents and humans) via sensors, it updates its internal SMM, which is synchronized with other agents. The agent uses its updated knowledge to adapt its behavior within the team and perform actions. The computational framework required to implement this vision consists of

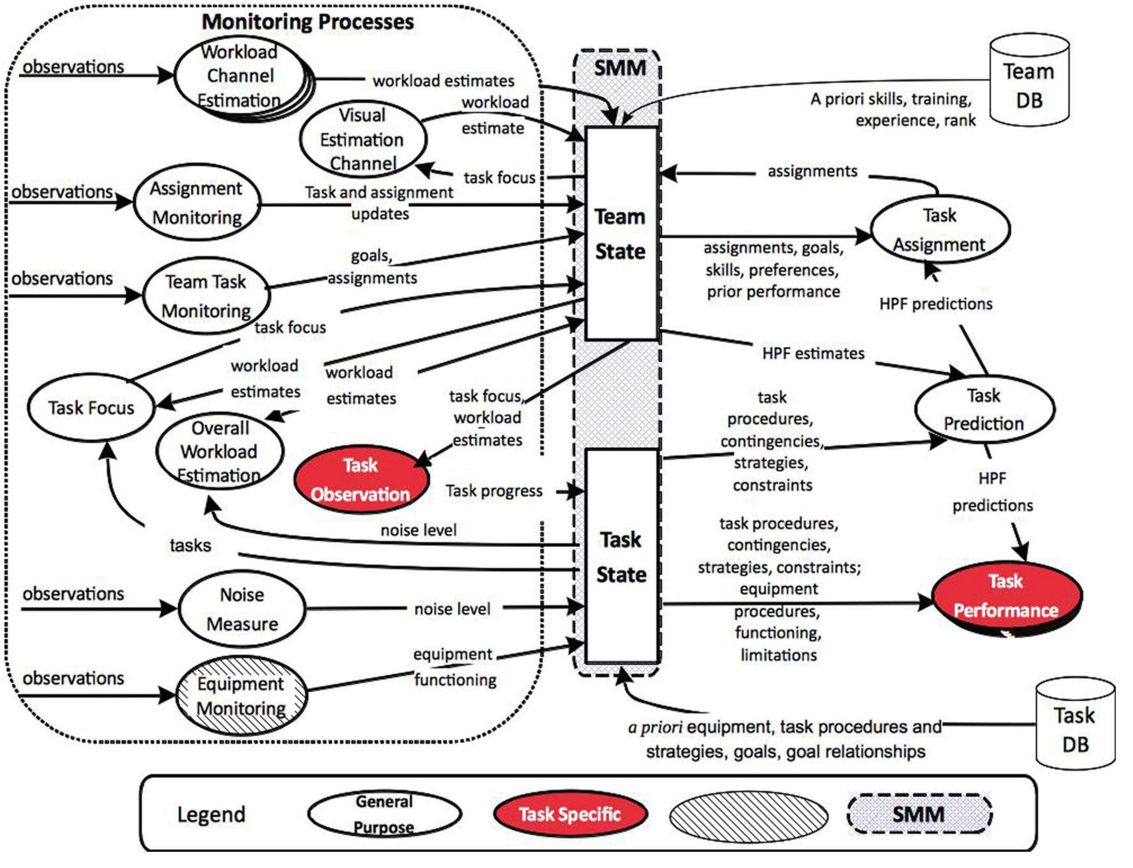

Figure 1. The shared mental model (SMM) computational framework. Ovals are processes, and arrows denote information flow. DB = database; HPF = human performance function.

a distributed data store, with processes to update SMM state information (left side of Figure 1) and control the agents based on the SMM data (right side). The computational framework needs to support the agent in (a) directly observing humans to monitor their behavior and estimate their current HPF values; (b) estimating HPFs, which are not ascertainable from direct monitoring; (c) tracking team/individual progress toward achieving goals; (d) predicting the effects of new or modified task assignments; (e) assigning new goals and tasks; and (f) adapting agent behavior based on the current state and future estimates. Steps 1–3 are update processes, whereas Steps 4–6 are control processes.

The team and task databases (see Figure 1) store a priori HPF values related to team members' skills, training, and experience (Table 2), as well as the team's tasks, equipment, and strategies (Table 1). This information is very stable and can be generated offline during training and via modeling. The databases are used to initialize the SMM and can be updated in real time with data from various update processes.

The computational framework is designed to support the use of tailorable and reusable components. The actual SMMs and databases can be reused and adapted to specific teams and tasks based on their contents. The underlying middleware for synchronizing the state components is also designed to be reusable across domains. These stable core components provide a highly reusable and adaptable interface for other components. Many of the general purpose processes, shown in white in Figure 1, can be reused and tailored to the specific domain and task via the contents of the SMM. For example, an HPF estimation process will use domain- and task-independent algorithms to make estimates, whereas only the SMM content is actually domain or task specific. Domain- and task-specific processes generally require tailoring the processes' algorithms to the domain and task. For instance, the equipment-monitoring process must be tailored to the specific types of equipment being monitored, whereas the task performance and augmentation algorithms can be tailored to specific domains and tasks, although they may use behavior specifications that can be executed via general purpose mechanisms.

Representing and Synchronizing SMMs

The goal of the computational framework is to capture the state information defined in the logical framework in a computationally efficient way. It is important to note that implementing a genuine SMM framework in artificial agents is different from simply using mental models in each agent, which occurs out of necessity in the humans; that is, humans use their mental models to realize SMMs by replicating the same knowledge in each (human) team member and by synchronizing the models via different forms of communication and observation. Clearly, this can be a cumbersome process, especially when only peer-to-peer communication is available, as every agent will have to be individually informed of an information update. SMMs in artificial agents do not have to replicate this constraint. Instead, an SMM can be implemented directly as a single distributed mental model shared among all artificial agents, which leads to more efficient representations and better synchronization with less communication overhead, especially among the artificial agents but also with the human team members.

Consider a subgroup G of a human team that receives an update that Room 7 collapsed: $C(r7)$. The members of the subgroup need to explicitly represent who knows and does not know about the event in each of their individual mental models, with the meta-belief that everyone in G knows. For example, the mental model of each team member i will have to include $C(r7)$, $Bel(H_j, C(r7))$ for all $j \neq i \in G$, $Bel(H_j, \neg C(r7))$ for $j \notin G$, and $Bel(H_j, Bel(H_k, C(r7)))$ for all $j \neq k \in G$. In contrast, a true SMM needs to add only $C(r7)$ and $Bel(H_j, \neg C(r7))$ for $j \notin G$, which all agents can access (it is assumed that if one agent believes a fact, all agents believe it). The smaller number of facts speeds up inferences, lowers redundancies in inferred facts, and reduces communication overhead with human members that cannot realize an SMM in the way that artificial agents can, as fewer facts need to be communicated to the humans about the agent's mental model (instead of communicating each fact for each agent individually, the facts can refer to the whole agent collective; e.g., "All agents believe that Room 7 collapsed").

The efficient shared aspect of SMMs, however, requires synchronization among agents. A distributed blackboard model (Corkill, 2003) can support asynchronous updates and distributed synchronization. During synchronization, conflict resolution follows the logical update rules for updating SMMs.

It is straightforward to capture logical primitives using object modeling (Rumbaugh, Blaha, Premerlani, Eddy, & Lorensen, 1991), which converts the logical entities into classes and logical predicates into associations. The Unified Modeling Language class models can capture these models; see Figure 2, which illustrates a simplified team state SMM. (A similar model is required for the task state.) Entities such as agents, roles, goals, and states are captured directly as classes, whereas predicates such as CAPABLE, KNOWS-HOW, and ADOPTED are captured as associations among the classes.

Some SMM state information—such as the team's goals, roles, agents (human and artificial), and tasks—is critical to any type of team activity and thus is part of any application's SMM state information. This basic state information can be augmented with additional information derived from the team's specific domain, objectives, and environment. State information pertaining to the human team members includes the current and predicated HPF values. As several processes within an agent asynchronously create, update, and use SMM state information, SMM state information has a variety of update requirements. Information such as task progress, current HPF values, and environmental conditions can be updated by continuous monitoring processes. Other information, such as new goals and assignments, are updated only when a triggering event occurs.

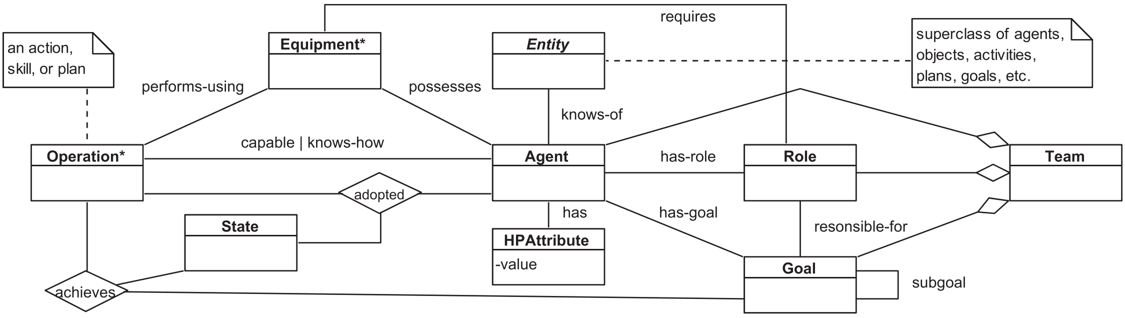

Figure 2. Simplified team state shared mental model. Asterisk denotes references to classes from the task state model. HP = human performance.

Updating SMMs

Update processes will revise the SMM state following the logical framework update rules and are extensible to changes in SMM information types. Some processes initialize the SMM with highly stable information (e.g., team member skills, training, experience, rank), whereas others will update SMM information in real time (e.g., cognitive workload, noise levels, current goals, and assignments). The real-time processes running on the robots will generally receive input directly from sensors, either located in the environment or attached to humans or robots.

Workload channel estimation. The workload channel estimation processes will calculate estimates for workload and the fluctuations over time. Each channel estimation process will accept metric observations and produce a value for that channel. Since workload can vary significantly, it is assessed in real time. Overall workload is derived from a number of submetrics (Boles, Bursk, Phillips, & Perdelwitz, 2007; Gawron, 2008). Workload is decomposed into the cognitive, auditory, visual, tactile, motor, and speech channels (Harriott et al., 2015), which is based on the Multiple Resource Questionnaire (Boles et al., 2007) and the IMPRINT Pro workload function (IMPRINT, 2012). The workload metrics in Table 3 do not adequately capture auditory, tactile, or motor demands. However, aspects of physical workload can be measured, allowing the tactile and motor channels to be replaced with a physical resource channel. The metrics can be captured with a bone microphone for the speech-based metrics and a physiological monitor, such as the BIOPAC® Systems Inc. Bioharness can capture the remaining metrics.

As workload increases, (a) the speaking rate; (b) the number of utterance repetitions and sentence fragments, false starts, and syntax errors; and (c) the number and duration of silence and filler pauses all increase (Berthold & Jameson, 1999), whereas utterance length decreases (Lively, Pisoni, Van Summers, & Bernacki, 1993). Training can mitigate the sensitivity of these metrics to individual differences (Hagmueller, Rank, & Kubin, 2006). The auditory resource can be assessed only via a speech response time metric. These metrics have to account for the ambient noise, provided by the noise measure process.

Real-time physiological metrics correlate to cognitive workload in that heart rate (Hankins & Wilson, 1998; Jorna, 1993; Roscoe, 1993) increases as workload increases (Castor et al., 2003), whereas heart rate variability (Aasman, Mulder, & Mulder, 1987; Harriott et al., 2015; Vicente, Thornton, & Moray, 1987) and respiration rate (Keller, Bless, Blomann, & Kleinbohl, 2001) decrease as cognitive workload increases (Aasman et al., 1987; Castor et al., 2003; Roscoe, 1992). An increase in galvanic skin response (i.e., the electrical conductance of the skin) indicates increased workload (Veltman & Gaillard, 1996), whereas skin temperature tends to decrease as workload increases (Collet & Averty, 2003).

The workload metrics will each be captured as one workload channel estimation process (Figure 1). All but the visual workload resource

TABLE 3: The Workload Metrics, Expected Change Under High-Workload Conditions, and Associated Channels

Metrics	Response	Overall	Cognitive	Auditory	Speech	Physical
Speaking rate	Increases	■	■		■	
No. of sentence fragments, false starts, and syntax errors	Increases	■	■		■	
No. of filler or delay utterances	Increases	■	■		■	
No. of utterance repetitions	Increases	■	■		■	
Response time: Speech	Increases	■	■	■	■	
Heart rate variability	Decreases	■	■			
Respiration rate	Decreases	■			■	■
Skin temperature	Increases	■				■

will be captured as a workload channel estimation process. It is infeasible to directly asses the visual resource channel; thus, it will be predicted according to models of the application scenario and knowledge of the task goals. The overall workload metric will be calculated after the visual channel demand has been predicted via the visual channel estimation process.

The workload channel estimation processes will produce continuous values (e.g., 0–100) for each channel; however, discrete values representing the continuous range will also be defined—for example, an extreme overload (85–100), a very high overload (70–84), a high overload (55–69), a moderate load (40–54), an underload (20–39), and an extreme underload (0–19). Additionally, the direction of change over time will be captured. Although measurements may remain within a range over time, the direction of change will help predict the human's future performance. Thus, a discrete value indicating increasing, decreasing, or stable workload will be determined and stored in the team state for use in behavior adaptation. Workload channel estimates will be stored in the team state, which will be used as inputs to the visual channel estimation, task focus, overall workload estimation, and task observation processes.

Visual Channel Estimation. The visual channel estimation process will estimate the visual resource channel workload based on the human's current task focus, skills, and rank within the organization, along with a priori information, such as visual workload models and subjective workload assessments from training. Tools such as IMPRINT Pro can be used to model the visual channel demand, which can be refined according to results from training and experience. The team database can be seeded with the model results, including gathered workload metrics and the human's skills and rank within the organization. Visual channel estimates will be stored in the team state and used by the overall workload estimate process.

Overall workload estimation. The overall workload estimation process will combine a human's workload and the visual channel estimates with the environmental noise measurements into a moving average of the human's overall workload. Each channel estimate will have custom weights derived from subjective workload metrics collected during training. The overall workload estimate with the individual channel estimates will be stored in the team state and used for behavior adaptation.

Noise measure. The noise measure process will measure the level of environmental noise in decibels via a sound-level meter or a dosimeter (Proctor & VanZandt, 2008) and output a discrete decibel range and the direction of change (increasing, decreasing or stable). Sustained loud noise levels can stress humans, increase workload, and negatively influence task progress, whereas sudden changes in noise levels can be distracting (Harris, 2011; Nassiri et al., 2013; Szalma & Hancock, 2011). The American Academy of Audiology provides seven discrete decibel noise levels (American Academy of Audiology, 2013), represented by four ranges:

<50, soft; 50–70, moderate; 70–90, loud; and, >90, very loud. Noise measurements will be stored in the task state and used by the overall workload estimation process to compute a moving average of the human's overall workload.

Task focus. Task focus will estimate on which tasks a human is currently focused, based on the human's HPFs and assigned goals and tasks. For example, an increased number of utterances may imply that a human is engaged in a specific task with a teammate, is providing task assignments or instructions, or is uncertain about a teammate's status. Thus, task focus will be estimated by identifying the subset of the human's assigned tasks whose expected HPF values most closely match the human's current channel estimates. Expected HPF values will be seeded in the task database. For instance, if a human has two tasks—(a) supervising a robot performing an area search and (b) searching an adjacent area—the task focus process will check each combination of tasks, {1}, {2}, and {1, 2}, to determine which subset most closely matches the current HPF estimates. Subsets will be ordered by likelihood, according to factors such as task priority, recently completed tasks, or human's prior location. Once a sufficiently close subset is found, it will be reported. The results of the task focus process will be stored in the team state and will be used by the visual channel estimation and task observation processes.

Task observation. Task observation will estimate a human's progress toward completing assigned tasks according to task information, current HPF estimates, the human's task focus, and a priori performance data. Although an agent can continually update the team on its progress, it is unreasonable to expect humans to do likewise. Thus, agents will estimate the human's progress. Although assessing a human's actions and ascertaining progress toward task achievement is difficult in general, a solution is feasible given an adequate model of the human's tasks and HPF estimates. For example, agents can use subtask structure to estimate progress. If a task requires a human to triage three victims, the agent can estimate that the task is two-thirds complete if the human has completed two of the triage subtasks. Additionally, agents can estimate progress using direct physical measurements (e.g., distance traveled) or a priori task performance data (e.g., average task completion time). The objective is to develop domain- and task-independent techniques; however, it is also necessary to understand the limitations of such approaches and when to augment them with domain- and task-specific techniques.

General update processes. The assignment-monitoring process will monitor interactions among teammates to determine when they need to be assigned or reassigned to roles or tasks and will update the SMM team state when such assignment occurs. The team task–monitoring processes will monitor team task status, including which tasks are active, achieved, or preceded, and update the SMM task state. Team task monitoring is straightforward, as demonstrated in our previous multiagent and robotic systems (DeLoach, 2009; Zhong & DeLoach, 2011). The equipment-monitoring processes will monitor equipment status and update the SMM task state. A priori equipment status will be loaded from the team database. Finally, an agent-learning process will monitor human and agent performance and update the knowledge, skills, abilities, preferences, and tendencies in the SMM team state.

Using SMMs to Adapt Agent Behavior

Control processes will use SMM state information to control agent and team behavior. These processes will cooperate to determine how and to what level to adapt agent behavior based on the human's current task assignments and expected performance changes. The task prediction process will estimate changes in human performance based on potential modifications to the human's current task assignments (i.e., the addition/removal of tasks). The task performance and augmentation process and task assignment process will use the predictions to allow agents to determine how to adapt the agent's behavior. The task performance and augmentation process will adapt internal behavior, whereas the task assignment process will adapt team behavior. For example, if human H_A is focused on triaging a victim, a robot working close by may tell H_B, the team leader, that it has completed its assigned search task. However, if H_A is the team leader, the robot may wait to inform H_A and take over one of H_A's lower-priority tasks.

The task prediction process will determine the expected changes in human performance based on potential modifications to the human's current task assignments (i.e., the addition/removal of tasks). The task prediction inputs will include (a) the human's current and proposed task assignments, (b) a set of the human's potential task assignment modifications, and (c) the current HPF estimates, including values for the individual workload channels. The process will produce the predicted overall workload and a prioritized list of workload channels for each potential modification. This process will consider each potential modification separately by predicting the effect on human performance.

The task performance and augmentation process will provide a mechanism to allow agents to modify their behavior to compensate for human performance degradation or external factors that negatively affect the team. For example, environmental noise can increase human stress and hinder communication. Additionally, humans may not have the appropriate training or skills to adequately perform an assigned task. This process will use a variety of inputs from the SMM for task-specific purposes, and the output will be appropriate agent behavior. Task augmentation will allow agents to assist other teammates by suggesting solutions, carrying out additional tasks, or modifying their own behavior to support their teammates. The task assignment process will consider the SMM state variables, including current task assignments and predictions. Using workload as the example, this process will determine the human's current state: no overload/underload—no changes are considered; extreme underload/overload—analyze task assignments and potential reallocations that significantly increase/decrease the human's overall workload, and allocate/reassign tasks; and underload/overload—agents may adapt their behavior or initiate a task reallocation to increase/decrease human workload. Tasks that can adjust the human's workload will be considered, with the modification effects determined by the task prediction process. The modifications will be based on their effect on human workload.

As an example, assume that human H_1 is assigned two goals: G_1 to triage victim V and G_2 to provide status information to the team leader. In addition, robot R_1 is assigned goal G_3 to assist in H_1's triage effort. As H_1 performs the triage process, sensor data indicate that H_1's heart rate variability is decreasing and the speaking rate and speech response time are increasing. The associated workload process estimation processes update H_1's values in the team state, changing H_1's cognitive and speech workloads from the *high* to *very high* state, thus pushing H_1's workload into the *extreme overload* state in the team state. The task assignment process monitors all the current assignments and, based on H_1's change to the *extreme overload* state, recomputes potential assignments for goal G_1. These computations are performed with an application-specific function, $a_{score}(agent,operation,goal)$, that accounts for basic agent capabilities, as modified by HPF values and the operations available to the agent to achieve the goal, similar to organizational assignment function (DeLoach, Oyenan, & Matson, 2008). If $a_{score}(R_1,o,G_2) > a_{score}(H_1,o',G_2)$ then, G_2 is reassigned from H_1 to R_1.

PRELIMINARY IMPLEMENTATIONS

Although there is currently no complete integrated implementation of the proposed architecture, various aspects of the system have been implemented and evaluated on robotic platforms and artificial agents in the context of team tasks. The organizational model for adaptive complex systems (DeLoach, 2009), for example, implements a subset of the proposed SMM focused on the assignment of agents to tasks within a team context. The basic model was used to implement an explosive device detection robot team (Zhong & DeLoach, 2011), which was extended to include physiological and subjective workload values and used to implement a hazardous materials reconnaissance scenario (Harriott et al., 2012).

Additional implementations of parts of the computational framework for SMMs (without monitoring of HPFs but with synchronization of SMMs across at least two robotic platforms) were implemented in the DIARC architecture

(Scheutz et al., 2013). For example, we demonstrated that pragmatic reasoning algorithms in the robot's natural language system can be used to automatically detect and correct wrong assumptions made by human teammates (e.g., Briggs & Scheutz, 2011). Conversely, human teammates can correct a robot's (wrong) beliefs explicitly through natural language instructions, which can lead to an update of the SMM for all robots in the team.

Two SMM implementation examples in the context of human-robot teams are presented. The examples indicate that full-fledged SMMs will significantly improve team performance. The first demonstrates how mental models figure critically in team-based human-robot interactions (in this case, natural language interactions), and the second demonstrates how robots can coordinate activities with human teammates through belief modeling and human-aware planning, even when communication is infeasible.

Simple SMMs

This example, which was implemented and evaluated on a robotic platform, presents a simple, minimal SMM that consists of several rules about perceptions, actions, and interactions (based on Briggs & Scheutz, 2012). The first rule about beliefs results from perceptions: If agent A_1 perceives another agent A_2 at location L, then A_1 also believes that A_2 is at that location:

$$\text{PERCEIVABLE}(A_1, at(A_2, L)) \Rightarrow \text{BELIEVES}(A_1, at(A_2, L)).$$

The following three rules focus on agent actions. If agent A has a goal to be at location L, then A is also going there:

$$\text{GOAL}(A, at(A, L)) \Rightarrow \text{GOINGTO}(A, L).$$

If agent A_1 is supposed to follow agent A_2 and if A_2 is heading to location L, A_1 is also going to L:

$$\text{FOLLOW}(A_1, A_2) \wedge \text{GOINGTO}(A_2, L) \Rightarrow \text{GOINGTO}(A_1, L).$$

If agent A_1 is supposed to meet with agent A_2 and if A_2 is currently at location L with no plans to move, then A_1 needs to move toward L:

$$\text{GOAL}(A_1, meet(A_1, A_2)) \wedge at(A_2, L) \wedge \neg\text{GOINGTO}(A_2, L') \Rightarrow \text{GOINGTO}(A_1, A_2), \text{(where } L' \neq L\text{)}.$$

The next rule triggers a notification event. If an agent A_1 is supposed to inform agent A_2 when a condition ϕ is achieved, then when ϕ is achieved, A_1 generates an intention-to-know ϕ for A_2, which can leverage the agent's dialogue-generation capabilities (generating a surface realization of the ϕ; for details, see Briggs & Scheutz, 2011):

$$\text{INFORM}(A_1, A_2, \phi) \wedge \phi \Rightarrow \text{ITK}(A_2, \phi).$$

Several rules were defined for communicative interactions and the interpretation of utterances—for example, if an agent A_1 commands another agent to go to a location L, then one can infer that A_2 will have a goal to be at L, that A_1 wants to be informed when A_2 reaches L, and that A_1 wants to know whether A_2 heard the command (the last inference is a dialogue behavior to generate appropriate acknowledgments):

$$\text{PROPOSES}(A_1, A_2, at(A_2, L)) \Rightarrow \text{GOAL}(A_2, at(A_2, L)) \wedge \text{INFORM}(A_1, A_2, at(A_2, L)) \wedge \text{ITK}(A_1, \text{HEARD}(A_2, \text{GOAL}(A_2, at(A_2, L)))).$$

Now suppose that H_1 orders robot R_1 to go to a triage location L_1 and R_2 to follow robot R_1. R_1 and R_2 share the SMM (and represent their beliefs about themselves in exactly the same way as they represent beliefs about other agents); thus, the SMM after the instruction contains the following items (we indicate an agent's beliefs in set notation, using B_A to denote the set of the agent's beliefs):

$$B_{H_1} := \{\text{G}(R_1, at(R_1, L_1)), \text{G}(R_2, follow(R_2, R_1))\},$$

$$B_{R_1} := \{\text{G}(R_1, at(R_1, L_1)), \text{inform}(H_1, at(R_1, L_1)), \text{ITK}(H_1, heard(R_1, G(R_1, at(R_1, L_1))))\},$$

$$B_{R_2} := \{\text{G}(R_2, follow(R_2, R_1)), inform(H_1, follow(R_2, R_1)), \text{ITK}(H_1, heard(R_2, G(R_2, follow(R_2, R_1))))\}.$$

The inform and ITK predicates are derived through the rule about what it means to receive an order. Suppose that H_1 wants to know where R_2 is going. R_2 can use the SMM to inspect the beliefs of all involved agents. R_2 can use the follow rules to infer that it is following R_1 and that R_1 is going to location L_1; thus, R_2 is going to location L_1 and answers H_2's question. This level of reasoning is infeasible without the SMM.

Coordination Through Human-Aware Planning

The robots in the prior example have an SMM, but shared models may not always be possible between agents and humans (e.g., the human is out of communications range or busy). Robots can be out of communication range, but the shared mental models are synchronized when connectivity is reestablished. An algorithm based on distributed truth maintenance can automatically resolve any inconsistencies due to the interrupted communication.

Suppose that human H_1 informs R_1 that triage support is needed at two triage locations, L_1 and L_2, H_2 is assigned to location L_2, and a medical kit is needed in each location. Moreover, R_1 and R_2 have medical kits, but R_1 is located closer to H_2 and R_2 is located very close to L_1. Thus, H_2 meets R_1 to pick up the medical kit, and R_2 proceeds to location L_1. However, assume that R_1 knows that H_2 believes that R_2 has a medical kit, but R_1 does not know whether H_2 also believes that R_1 has a medical kit. In this case, R_1 does not plan to meet H_2 but instead proceeds directly to location L_1 because H_2 will likely move toward R_2 and not R_1; furthermore, there is no way to contact H_2 with updated information about the robots' locations. R_1 can determine this fact by generating all reasonable plans for H_2 based on the information available in the SMM shared by R_1 and R_2 about H_2's beliefs and goals (for details on this situation and the associated planning, see Talamadupula, Briggs, Chakraborti, Scheutz, & Kambhampati, 2014). Once R_1 has determined that H_2 intends to meet up with R_2, it adopts the goal to go to location L_1, and R_2 adopts the goal to move toward H_2, thus reducing H_2's effort while optimizing the overall team's performance (by delivering the medical kits more quickly to each triage location).

CONCLUSIONS

The literature has demonstrated the important role that mental models and SMMs play in improving human team performance. As technology improves and we move toward a world that will team various types of autonomous (or semiautonomous) agents with humans, one must consider the potential agent capabilities. Such human-agent teams will exist and become prevalent only if they are able to carry out tasks as effectively and with as good as or better performance than human-only teams. This paper introduces a comprehensive framework for defining, implementing, and applying SMMs in artificial agents such as robots to improve the performance of mixed human-agent teams. Specifically, the general formal framework provides the representational and formal principles for capturing important state information and knowledge-based aspects of SMMs, whereas the computational framework provides a detailed overview of a computational architecture and the various processes and their function needed to create, update, and maintain SMMs. The prototypical SMM scenario illustrates how the SMM framework can be applied to a real-world situation, including how various framework components will arise in artificial agents.

Although artificial intelligence, robotics, and human-robot interaction researchers have worked toward designing mixed human-agent (robot) teams incorporating principles such as belief-desires-intention frameworks, theory-of-mind representations, coordination rules and mechanisms, and task assignment algorithms, there is currently no comprehensive computational SMM framework that addresses the breadth and depth of the presented computational framework, let alone one that has been implemented for artificial agent systems.

REFERENCES

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors*, *29*(2), 161–170.

Adams, J. A., Harriott, C. E., Zhuang, R., & DeLoach, S. A. (2012). *Human-robot teams informed by human performance moderator functions: A final report* (Technical Report No. FA9550-09-1-0108). Wright-Patterson Air Force Base, OH: Air Force Office of Scientific Research.

American Academy of Audiology. (2013). *Levels of noise in decibels*. Retrieved from http://www.audiology.org/resources/brochures/Pages/default.aspx

Banks, A. P., & Millward, L. J. (2000). Running shared mental models as a distributed cognitive process. *British Journal of Psychology*, *91*, 513–531.

Berthold, A., & Jameson, A. (1999). Interpreting symptoms of cognitive load in speech input. In *Proceedings of the International Conference on User Modeling* (pp. 235–244).

Boles, D., Bursk, J., Phillips, J., & Perdelwitz, J. (2007). Predicting dual-task performance with the multiple resources questionnaire (MRQ). *Human Factors*, *49*(1), 32–45.

Borgman, C. L. (1986). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *Journal of Man-Machine Studies*, *24*, 47–64.

Bosse, T., Memon, Z., & Treur, J. (2007). A two-level BDI-agent model for theory of mind and its use in social manipulation. In *Proceedings of the Artificial and Ambient Intelligence Conference* (pp. 335–342).

Bridgeland, D. M., & Huhns, M. N. (1990). Distributed truth maintenance. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 72–77).

Briggs, G., & Scheutz, M. (2011). Facilitating mental modeling in collaborative human-robot interaction through adverbial cues. In *Proceedings of the Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 239–247).

Briggs, G., & Scheutz, M. (2012). Multi-modal belief updates in multi-robot human-robot dialogue interactions. In *Proceedings of the AISB/IACAP World Congress*.

Burkhard, H.-D., Bach, J., Berger, R., Brunswieck, B., & Gollin, M. (2002). Mental models for robot control. *Advances in Plan-Based Control of Robotic Agents, Lecture Notes in Computer Science*, *2466*, 71–88.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Current issues in individual and team decision making* (pp. 221–246). Hillsdale, NJ: Erlbaum.

Castelfranchi, C. (1998). Modeling social action for AI agents. *Artificial Intelligence*, *103*, 157–182.

Castor, M., Hanson, E., Svensson, E., Nahlinder, S., LeBlaye, P., MacLeod, I., . . . Berggren, P. (2003). *GARTEUR handbook of mental workload measurement: Technical chapter 145* (Technical Report No. FM AG13: 164 edition). Bedfordshire, UK: Group for Aeronautical Research and Technology in Europe.

Clark, H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Collet, C., & Averty, P. (2003). Subjective aspects of mental workload in air-traffic control. In G. Hockey, A. Gaillard, & O. Burov (Eds.), *Operator functional state* (pp. 291–302). Amsterdam, Netherlands: IOS Press.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C. A., & Cannon-Bowers, J. A. (2003). Measuring team knowledge. *Group Dynamics: Theory, Research, and Practice*, *7*(3), 179–199.

Corkill, D. D. (2003). Collaborating software: Blackboard and multi-agent systems & the future. In *Proceedings of the International Lisp Conference* (Vol. 10).

Dahl, T. S., Matarić, M. J., & Sukhatme, G. S. (2009). Multi-robot task allocation through vacancy chain scheduling. *Journal of Robotics and Autonomous Systems*, *57*(6), 674–687.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A metaanalysis. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 1–14.

DeLoach, S. (2009). OMACS: A framework for adaptive, complex systems. In V. Dignum (Ed.), *Handbook of research on multi-agent systems: Semantics and dynamics of organizational models* (pp. 76–104). Hershey, PA: IGI Global.

DeLoach, S., Oyenan, W. H., & Matson, E. T. (2008). A capabilities-based model for adaptive organizations. *Autonomous Agents and Multi-Agent Systems*, *16*(1), 13–56.

DuRussell, L. A., & Derry, S. J. (2005). Schema (mis)alignment in interdisciplinary teamwork. In S. J. Derry, C. D. Schunn, & M. A. Gernsbacher (Eds.), *Interdisciplinary collaboration: An emerging cognitive science* (pp. 187–220). London, UK: Erlbaum.

Espevik, R., Johnsen, B. H., & Eid, J. (2011). Outcomes of shared mental models of team members in cross training and high-intensity simulations. *Journal of Cognitive Engineering and Decision Making*, *5*(4), 352–377.

Fan, X., & Yen, J. (2007). Realistic cognitive load modeling for enhancing shared mental models in human-agent collaboration. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 395–402).

Fan, X., & Yen, J. (2011). Modeling cognitive loads for evolving shared mental models in human-agent collaboration. *IEEE Transactions on Systems, Man and Cybernetics–Part B: Cybernetics*, *41*(2), 354–367.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*, 143–166.

Gawron, V. (2008). *Human performance, workload, and situational awareness measures handbook*. Boca Raton, FL: CRC Press.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (2003). The belief-desire-intention model of agency. In J. P. Müller, A. S. Rao, & M. P. Singh (Eds.), *Lecture notes in computer science: Intelligent agents V: Agents theories, architectures, and languages* (Vol. 1555, pp. 1–10). Berlin, Germany: Springer.

Goodrich, M. A., & Yi, D. (2013). Toward task-based mental models of human-robot teaming: A Bayesian approach. In R. Shumaker (Ed.), *Virtual augmented and mixed reality: Designing and developing augmented and virtual environments* (Vol. 8021, pp. 267–276). Heidelberg, Germany: Springer Berlin.

Gray, S. H. (1990). Using protocol analyses and drawings to study mental model construction during hypertext navigation. *International Journal of Human-Computer Interaction*, *2*, 359–377.

Hagmueller, M., Rank, E., & Kubin, G. (2006). *Evaluation of the human voice for indications of workload induced stress in the aviation environment* (ECC Note No. 18/06). Brussels, Belgium: European Organisation for the Safety of Air Navigation.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation Space and Environmental Medicine*, *69*(4), 360–367.

Harriott, C. E., & Adams, J. A. (2013). Modeling human performance for human-robot systems. In D. B. Kaber (Ed.), *Reviews in human factors and ergonomics: Human performance in teleoperations and beyond* (Vol. 9, pp. 94–103). Thousand Oaks, CA: SAGE.

Harriott, C. E., Buford, G., Zhang, T., & Adams, J. A. (2015). Mental workload and task performance in peer-based human-robot teams. *Journal of Human-Robot Interaction*, *14*(2), 61–96.

Harriott, C. E., Zhang, T., & Adams, J. A. (2013). Assessing physical workload for human-robot peerbased teams. *International Journal of Human-Computer Studies*, *71*(7–8), 821–837.

Harriott, C. E., Zhuang, R., Adams, J. A., & DeLoach, S. (2012). Towards using human performance moderator functions in human-robot teams. In *Proceedings of International Workshop on Human-Agent Interaction Design and Models at the International Conference on Autonomous Agents and Multiagent Systems*.

Harris, D. (Ed.). (2011). *Human performance on the flight deck*. Surrey, UK: Ashgate.

Held, C., Knaff, M., & Vosgerau, G. (Eds.). (2006). *Mental models and the mind: Current developments in cognitive psychology, neuroscience, and philosophy of the mind*. Amsterdam, Netherlands: Elsevier.

Humphrey, C. M. (2009). *Information abstraction visualization for human-robot interaction* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.

Humphrey, C. M., & Adams, J. A. (2009). Robotic tasks for CBRNE incident response. *Advanced Robotics*, *23*, 1217–1232.

Humphrey, C. M., & Adams, J. A. (2011). Analysis of complex, team-based systems: Augmentations to goal-directed task analysis and cognitive work analysis. *Theoretical Issues in Ergonomics*, *12*(2), 149–175.

Humphrey, C. M., & Adams, J. A. (2013). Cognitive information flow analysis. *Cognition, Technology and Work*, *15*(2), 133–152.

IMPRINT. (2012). *IMPRINT pro user guide: Volumes 1–3*. http://www.arl.army.mil/www/pages/446/IMPRINTProvol1.pdf .

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, UK: Cambridge University Press.

Jonker, C., Riemsdijk, M., & Vermeulen, B. (2011). Shared mental models. In M. Vos, N. Fornara, J. Pitt, & G. Vouros (Eds.), *Coordination, organizations, institutions, and norms in agent systems VI* (Vol. 6541, pp. 132–151). Heidelberg, Germany: Springer Berlin.

Jonker, C., van Riemsdijk, M. B., & Vermeulen, B. (2010). Shared mental models: A conceptual analysis. In *Proceedings of the International Conference on Coordination, Organizations, Institutions, and Norms in Agent Systems* (pp. 132–151).

Jorna, P. (1993). Heart-rate and workload variations in actual and simulated flight. *Ergonomics*, *36*(9), 1034–1054.

Keisler, S., & Goetz, J. (2002). Mental models and cooperation with robotic assistants. In *Proceedings of the ACM Conference on Human Factors in Computer Systems* (pp. 576–577).

Keller, J., Bless, H., Blomann, F., & Kleinbohl, D. (2001). Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary control. *Journal of Experimental Social Psychology*, *47*(4), 849–852.

Kennedy, W. G., & Trafton, J. G. (2007). Using simulations to model shared mental models. In *Proceedings of the International Conference on Cognitive Modeling* (pp. 253–254).

Lee, M., Johnson, T., & Jin, M. (2012). Toward understanding the dynamic relationship between team and task shared mental models as determinants of team and individual performances. *International Journal of Information Technology and Business Management*, *8*(1), 1–14.

Lee, M., Johnson, T., Lee, Y., O'Connor, D., & Khalil, M. (2004). *The conceptual framework of factors affecting shared mental model*. Bloomington, IN: Association for Educational Communications and Technology.

Lee, S.-L., Keisler, S., Lau, I. Y.-M., & Chiu, C.-Y. (2005). Human mental models of humanoid robots. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 2767–2772).

Libiere, C., Jentsch, F., & Ososky, S. (2013). Cognitive models of decision making processes for humanrobot interaction. In *Proceedings of the International Conference on Virtual Augmented and Mixed Reality, Lecture Notes in Computer Science* (Vol. 8021, p. 285–294).

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*, 403–418.

Lively, S., Pisoni, D., Van Summers, W., & Bernacki, R. (1993). Effects of cognitive workload on speech production: Acoustic analysis and perceptual consequences. *Journal of the Acoustical Society of America*, *93*(5), 2962–2973.

Luo, L., Chakraborty, N., & Sycara, K. (2015). Distributed algorithms for multi-robot task assignment with task deadline constraints. *IEEE Transactions on Automation Sciences and Engineering*, *12*(3), 876–888.

Mathieu, J. E., Goodwin, G. F., Heffner, T. S., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283.

Mathieu, J. E., Rapp, T. L., Maynard, M. T., & Mangos, P. M. (2009). Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Human Performance*, *23*(1), 22–40.

Miwa, H., Okuchi, T., Itoh, K., Takanobu, H., & Takanishi, A. (2003). A new mental model for humanoid robots for human friendly communication - Introduction of learning system, mood vector and second order equations of emotion. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 3588–3593).

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, *36*(4), 876–910.

Mumaw, R. J., Roth, E. M., Vicenti, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors*, *42*(1), 36–55.

Nassiri, P., Monazan, M., Dehaghi, B. F., Ibrahimi Ghavan Adadi, L., Zakerian, S. A., & Azam, K. (2013). The effect of noise on human performance: A clinical trial. *International Journal of Occupational and Environmental Medicine*, *4*(2), 87–95.

Neerincx, M. A., de Greef, T., Smets, N. J. J. M., & Sam, M. P. (2011). Shared mental models of distributed human-robot teams for coordinated disaster responses. In *Proceedings of the AAAI Fall Symposium* (FS-11-05).

Nikolaidis, S., & Shah, J. A. (2013). Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 33–40).

Ososky, S., Schuster, D., Jentsch, F., Fiore, S., Shumaker, R., Lebiere, C., . . . Stentz, A. (2012). The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In *Proceedings of SPIE Unmanned Systems Technology XIV* (Vol. 8387, p. 10–1).

Park, K.-S., Kwon, D.-S., & Park, M. (2007). A design of the mental model of a cognitive robot. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication* (pp. 905–911).

Proctor, R. W., & VanZandt, T. (2008). *Human factors in simple and complex systems*. Boca Raton, FL: CRC Press.

Richards, D. (2001). Coordination and shared mental models. *American Journal of Political Science*, *45*(2), 259–276.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*, 259–287.

Roscoe, A. H. (1993). Heart-rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, *36*(9), 223–228.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*, 349–363.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. E. (1991). *Object-oriented modeling and design*. Englewood Cliffs, NJ: Prentice Hall.

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies*, *55*, 85–107.

Scheutz, M., Briggs, G., Cantrell, R., Krause, E., Williams, T., & Veale, R. (2013). Novel mechanisms for natural human-robot interactions in the DIARC architecture. In *Proceedings of AAAI Workshop on Intelligent Robotic Systems*.

Scholtz, J. (2003). Theory and evolution of human robot interactions. In *Proceedings of IEEE International Conference on System Sciences* (pp. 125–134).

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: Part I. Enabling science with PMFServ. *Presence: Teleoperators and Virtual Environments*, *15*(2), 139–162.

Stubbs, K., Wettergreen, D., & Hinds, P. J. (2007). Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems*, *22*(2), 42–50.

Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L., & Otero, N. R. (2010). Exploring human mental models of robots through explicitation interviews. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication* (pp. 638–645).

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682–707.

Talamadupula, K., Briggs, G., Chakraborti, T., Scheutz, M., & Kambhampati, S. (2014). Coordination in human-robot teams using mental modeling and plan recognition. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2957–2962).

Tambe, M. (1997). Towards flexible teamwork. *Journal of Artificial Intelligence Research*, *7*, 83–124.

Veltman, J., & Gaillard, A. (1996). *Physiological workload reactions to increasing levels of task difficulty*. (Technical Report No. TM-96-A026). The Hague, Netherlands: TNO Human Factors Research Institute.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors*, *29*(2), 171–182.

Weaver, R., Silverman, B. G., Shin, H., & Dubois, R. (2001). Modeling and simulating terrorist decision-making: A performance moderator function approach to generating virtual opponents. In *Proceedings of the Conference on Computer Generated Forces and Behavioral Representation* (pp. 39–44).

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. E. (2003). *Introduction to human factors engineering* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. *Human Factors*, *49*(2), 243–256.

Zhang, Y. (2008). Role-based shared mental models. In *Proceedings of the International Symposium on Collaborative Technologies and Systems* (pp. 424–431).

Zhong, C., & DeLoach, S. A. (2011). Runtime models for automatic reorganization of multi-robot systems. In *Proceedings of the International Symposium on Software Engineering for Adaptive and Self-Managing Systems* (pp. 20–29).

Matthias Scheutz is a professor in cognitive and computer science in the Department of Computer Science at Tufts University. He earned a PhD in philosophy from the University of Vienna in 1995 and a joint PhD in cognitive science and computer science from Indiana University, Bloomington, in 1999. He has more than 250 peer-reviewed publications in artificial intelligence, natural language processing, cognitive modeling, robotics, and human-robot interaction. His current research focuses on complex cognitive robots with natural language capabilities.

Scott A. DeLoach is a professor and the head of the Department of Computer Science at Kansas State University. He has been on faculty since 2001, prior to which he was an officer in the U.S. Air Force for 20 years. Scott earned his PhD from the Air Force Institute of Technology in 1996 and has over 90 peer-reviewed publications in software engineering, multiagent systems, artificial intelligence, robotics, and human-robot interaction. His current research interests focus on applying software engineering methods, techniques, and models to the design and development of intelligent, complex, adaptive, and autonomous systems.

Julie A. Adams is a professor of computer science and computer engineering in the Department of Electrical Engineering and Computer Science at Vanderbilt University. She earned a PhD in computing and information sciences at the University of Pennsylvania in 1995 and previously worked in human factors at Honeywell, Inc. and the Eastman Kodak Company. She has more than 150 peer-reviewed publications in artificial intelligence, robotics, human-computer interaction, and human-robot interaction. Her current research interests focus on complex human-systems interaction and distributed artificial intelligence for multiple robot systems.