Realistic Cognitive Load Modeling for Enhancing Shared Mental Models in Human-Agent Collaboration

Xiaocong Fan
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802
zfan@ist.psu.edu

John Yen
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802
jyen@ist.psu.edu

ABSTRACT

Human team members often develop shared expectations to predict each other's needs and coordinate their behaviors. In this paper the concept "Shared Belief Map" is proposed as a basis for developing realistic shared expectations among a team of Human-Agent-Pairs (HAPs). The establishment of shared belief maps relies on inter-agent information sharing, the effectiveness of which highly depends on agents' processing loads and the instantaneous cognitive loads of their human partners. We investigate HMM-based cognitive load models to facilitate team members to "share the right information with the right party at the right time". The shared belief map concept and the cognitive/processing load models have been implemented in a cognitive agent architecture—SMMall. A series of experiments were conducted to evaluate the concept, the models, and their impacts on the evolving of shared mental models of HAP teams.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*

General Terms

Design, Experimentation, Human Factors

Keywords

Cognitive modeling, Human-centered teamwork, Shared belief maps, Multi-party communication

1. INTRODUCTION

The entire movement of "agent paradigm" was spawned, at least in part, by the perceived importance of fostering human-like adjustable autonomy. Human-centered multi-agent teamwork has thus attracted increasing attentions in multi-agent systems field [2, 10, 4]. Humans and autonomous systems (agents) are generally thought to be complementary: while humans are limited by their cognitive capacity in information processing, they are superior in spatial, heuristic, and analogical reasoning; autonomous systems can continuously learn expertise and tacit problem-solving knowledge from humans to improve system performance. In short, humans and agents can team together to achieve better performance, given that they could establish certain mutual awareness to coordinate their mixed-initiative activities.

However, the foundation of human-agent collaboration keeps being challenged because of nonrealistic modeling of mutual awareness of the state of affairs. In particular, few researchers look beyond to assess the principles of modeling shared mental constructs between a human and his/her assisting agent. Moreover, human-agent relationships can go beyond partners to teams. Many informational processing limitations of individuals can be alleviated by having a group perform tasks. Although groups also can create additional costs centered on communication, resolution of conflict, and social acceptance, it is suggested that such limitations can be overcome if people have *shared cognitive structures* for interpreting task and social requirements [8]. Therefore, there is a clear demand for investigations to broaden and deepen our understanding on the principles of shared mental modeling among members of a mixed human-agent team.

There are lines of research on multi-agent teamwork, both theoretically and empirically. For instance, Joint Intention [3] and SharedPlans [5] are two theoretical frameworks for specifying agent collaborations. One of the drawbacks is that, although both have a deep philosophical and cognitive root, they do not accommodate the modeling of human team members. Cognitive studies suggested that teams which have shared mental models are expected to have common expectations of the task and team, which allow them to predict the behavior and resource needs of team members more accurately [14, 6]. Cannon-Bowers et al. [14] explicitly argue that team members should hold compatible models that lead to common "expectations". We agree on this and believe that the establishment of *shared expectations* among human and agent team members is a critical step to advance human-centered teamwork research.

It has to be noted that the concept of shared expectation can broadly include role assignment and its dynamics, teamwork schemas and progresses, communication patterns and intentions, etc. While the long-term goal of our research is to understand how shared cognitive structures can enhance human-agent team performance, the specific objective of the work reported here is to develop a computational cognitive

AAMAS'07 May 14–18, 2007, Honolulu, Hawai'i, USA.

capacity model to facilitate the establishment of shared expectations. In particular, we argue that to favor human-agent collaboration, an agent system should be designed to allow the estimation and prediction of human teammates' (relative) cognitive loads, and use that to offer improvised, unintrusive help. Ideally, being able to predict the cognitive/processing capacity curves of teammates could allow a team member to help *the right party at the right time*, avoiding unbalanced work/cognitive loads among the team.

The last point is on the modeling itself. Although an agent's cognitive model of its human peer is not necessarily to be descriptively accurate, having at least a realistic model can be beneficial in offering unintrusive help, bias reduction, as well as trustable and self-adjustable autonomy. For example, although humans' use of cognitive simplification mechanisms (e.g., heuristics) does not always lead to errors in judgment, it can lead to predictable biases in responses [8]. It is feasible to develop agents as cognitive aids to alleviate humans' biases, as long as an agent can be trained to obtain a model of a human's cognitive inclination. With a realistic human cognitive model, an agent can also better adjust its automation level. When its human peer is becoming overloaded, an agent can take over resource-consuming tasks, shifting the human's limited cognitive resources to tasks where a human's role is indispensable. When its human peer is underloaded, an agent can take the chance to observe the human's operations to refine its cognitive model of the human. Many studies have documented that human choices and behaviors do not agree with predictions from rational models. If agents could make recommendations in ways that humans appreciate, it would be easier to establish trust relationships between agents and humans; this in turn, will encourage humans' automation uses.

The rest of the paper is organized as follows. In Section 2 we review cognitive load theories and measurements. A HMM-based cognitive load model is given in Section 3 to support resource-bounded teamwork among human-agent-pairs. Section 4 describes the key concept "shared belief map" as implemented in SMMall, and Section 5 reports the experiments for evaluating the cognitive models and their impacts on the evolving of shared mental models.

2. COGNITIVE CAPACITY–OVERVIEW

People are information processors. Most cognitive scientists [8] believe that human information-processing system consists of an executive component and three main information stores: (a) sensory store, which receives and retains information for one second or so; (b) working (or short-term) memory, which refers to the limited capacity to hold (approximately seven elements at any one time [9]), retain (for several seconds), and manipulate (two or three information elements simultaneously) information; and (c) long-term memory, which has virtually unlimited capacity [1] and contains a huge amount of accumulated knowledge organized as schemata. Cognitive load studies are, by and large, concerned about working memory capacity and how to circumvent its limitations in human problem-solving activities such as learning and decision making.

According to the cognitive load theory [11], *cognitive load* is defined as a multidimensional construct representing the load that a particular task imposes on the performer. It has a causal dimension including causal factors that can be characteristics of the subject (e.g. expertise level), the task (e.g. task complexity, time pressure), the environment (e.g. noise), and their mutual relations. It also has an assessment dimension reflecting the measurable concepts of mental load (imposed exclusively by the task and environmental demands), mental effort (the cognitive capacity actually allocated to the task), and performance.

Lang's information-processing model [7] consists of three major processes: encoding, storage, and retrieval. The encoding process selectively maps messages in sensory stores that are relevant to a person's goals into working memory; the storage process consolidates the newly encoded information into chunks, and form associations and schema to facilitate subsequent recalls; the retrieval process searches the associated memory network for a specific element/schema and reactivates it into working memory. The model suggests that processing resources (cognitive capacity) are independently allocated to the three processes. In addition, working memory is used both for holding and for processing information [1]. Due to limited capacity, when greater effort is required to process information, less capacity remains for the storage of information. Hence, the allocation of the limited cognitive resources has to be balanced in order to enhance human performance. This comes to the issue of measuring cognitive load, which has proven difficult for cognitive scientists.

Cognitive load can be assessed by measuring mental load, mental effort, and performance using rating scales, psychophysiological (e.g. measures of heart activity, brain activity, eye activity), and secondary task techniques [12]. Self-ratings may appear questionable and restricted, especially when instantaneous load needs to be measured over time. Although physiological measures are sometimes highly sensitive for tracking fluctuating levels of cognitive load, costs and work place conditions often favor task- and performance-based techniques, which involve the measure of a secondary task as well as the primary task under consideration. Secondary task techniques are based on the assumption that performance on a secondary task reflects the level of cognitive load imposed by a primary task [15]. From the resource allocation perspective, assuming a fixed cognitive capacity, any increase in cognitive resources required by the primary task must inevitably decrease resources available for the secondary task [7]. Consequently, performance in a secondary task deteriorates as the difficulty or priority of the primary task increases. The level of cognitive load can thus be manifested by the secondary task performance: the subject is getting overloaded if the secondary task performance drops.

A secondary task can be as simple as detecting a visual or auditory signal but requires sustained attention. Its performance can be measured in terms of reaction time, accuracy, and error rate. However, one important drawback of secondary task performance, as noted by Paas [12], is that it can interfere considerably with the primary task (competing for limited capacity), especially when the primary task is complex. To better understand and measure cognitive load, Xie and Salvendy [16] introduced a conceptual framework, which distinguishes instantaneous load, peak load, accumulated load, average load, and overall load. It seems that the notation of instantaneous load, which represents the dynamics of cognitive load over time, is especially useful for monitoring the fluctuation trend so that free capacity can be exploited at the most appropriate time to enhance the overall performance in human-agent collaborations.

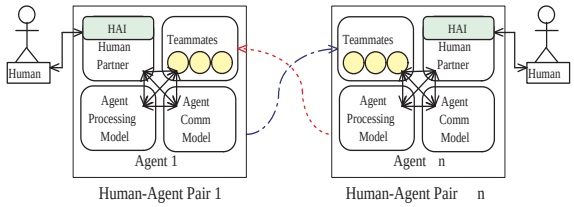
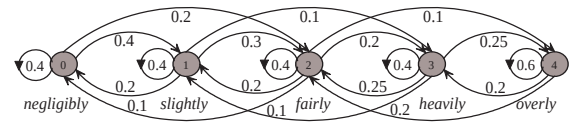


Figure 1: Human-centered teamwork model.



$$B = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \geq 9 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.02 & 0.03 & 0.05 & 0.1 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 0.05 & 0.05 & 0.1 & 0.7 & 0.1 \\ 0 & 0 & 0 & 0 & 0.01 & 0.02 & 0.45 & 0.4 & 0.1 & 0.02 \\ 0.02 & 0.03 & 0.05 & 0.15 & 0.4 & 0.3 & 0.03 & 0.02 & 0 & 0 \\ 0.1 & 0.3 & 0.3 & 0.2 & 0.1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Figure 2: A HMM Cognitive Load Model.

3. HUMAN-CENTERED TEAMWORK MODEL

People are limited information processors, and so are intelligent agent systems; this is especially true when they act under hard or soft timing constraints imposed by the domain problems. In respect to our goal to build realistic expectations among teammates, we take two important steps.

First, agents are resource-bounded; their processing capacity is limited by computing resources, inference knowledge, concurrent tasking capability, etc. We withdraw the assumption that an agent knows *all* the information/intentions communicated from other teammates. Instead, we contend that due to limited processing capacity, an agent may only have opportunities to process (make sense of) a portion of the incoming information, with the rest ignored. Taking this approach will largely change the way in which an agent views (models) the involvement and cooperativeness of its teammates in a team activity. In other words, the establishment of shared mental models regarding team members' beliefs, intentions, and responsibilities can no longer rely on inter-agent communication only. This being said, we are not dropping the assumption that "teammates are trustable". We still stick to this, only that team members cannot overtrust each other; an agent has to consider the *possibility* that its information being shared with others might not be as effective as expected due to the recipients' limited processing capacities. Second, human teammates are bounded by their cognitive capacities. As far as we know, the research reported here is the first attempt in the area of human-centered multi-agent teamwork that really considers building and using human's cognitive load model to facilitate teamwork involving both humans and agents.

We use $\langle H_i, A_i \rangle$ to denote Human-Agent-Pair (HAP) i.

3.1 Computational Cognitive Capacity Model

An intelligent agent being a cognitive aid, it is desirable that the model of its human partner implemented within the agent is *cognitively-acceptable*, if not descriptively accurate. Of course, building a cognitive load model that is cognitively-acceptable is not trivial; there exist a variety of cognitive load theories and different measuring techniques. We here choose to focus on the performance variables of secondary tasks, given the ample evidence supporting secondary task performance as a highly sensitive and reliable technique for measuring human's cognitive load [12]. It's worth noting that just for the purpose of estimating a human subject's cognitive load, any artificial task (e.g, pressing a button in response to unpredictable stimuli) can be used as a secondary task to force the subject to go through. However, in a realistic application, we have to make sure that the selected secondary task interacts with the primary task in meaningful ways, which is not easy and often depends on the domain problem at hand. For example, in the experiment below, we used the number of newly available information correctly recalled as the secondary task, and the effectiveness of information sharing as the primary task. This is realistic to intelligence workers because in time stress situations they have to know what information to share in order to effectively establish an awareness of the global picture.

In the following, we adopt the Hidden Markov Model (HMM) approach [13] to model human's cognitive capacity. It is thus assumed that at each time step the secondary task performance of a human subject in a team composed of human-agent-pairs (HAP) is observable to all the team members. Human team members' secondary task performance is used for estimating their hidden cognitive loads.

A HMM is denoted by $\lambda = \langle N, V, A, B, \pi \rangle$, where N is a set of hidden states, V is a set of observation symbols, A is a set of state transition probability distributions, B is a set of observation symbol probability distributions (one for each hidden state), and π is the initial state distribution.

We consider a 5-state HMM model of human cognitive load as follows (Figure 2). The hidden states are 0 (negligibly-loaded), 1 (slightly-loaded), 2 (fairly-loaded), 3 (heavily-loaded), and 4 (overly loaded). The observable states are tied with secondary task performance, which, in this study, is measured in terms of the number of items correctly recalled. According to Miller's 7 ± 2 rule, the observable states take integer values from 0 to 9 (the state is 9 when the number of items correctly recalled is no less than 9). For the example B Matrix given in Fig. 2, it is very likely that the cognitive load of the subject is "negligibly" when the number of items correctly recalled is larger than 9.

However, to determine the current hidden load status of a human partner is not trivial. The model might be oversensitive if we only consider the last-step secondary task performance to locate the most likely hidden state. There is ample evidence suggesting that human cognitive load is a continuous function over time and does not manifest sudden shifts unless there is a fundamental changes in tasking demands. To address this issue, we place a constraint on the state transition coefficients: no jumps of more than 2 states are allowed. In addition, we take the position that, a human subject is very likely overloaded if his secondary task performance is mostly low in recent time steps, while he is very likely not overloaded if his secondary task performance is mostly high recently. This leads to the following Windowed-HMM approach.

Given a pre-trained HMM λ of human cognitive load and the recent observation sequence O_t of length w, let parameter w be the effective window size, ε_t^λ be the *estimated* hidden state at time step t. First apply the HMM to the observation sequence to find the optimal sequence of hidden states $S_t^\lambda = s_1 s_2 \cdots s_w$ (Viterbi algorithm). Then, compute the *estimated* hidden state ε_t^λ for the current time step, viewing it as a function of S_t^λ. We consider all the hidden states in S_t^λ, weighted by their respective distance to $\varepsilon_{t-1}^\lambda$ (the estimated state of the last step): the closer of a state in S_t^λ

to $\varepsilon_{t-1}^{\lambda}$, the higher probability of the state being ε_t^{λ}. ε_t^{λ} is set to be the state with the highest probability (note that a state may have multiple appearances in S_t^{λ}). More formally, the probability of state $s \in S$ being ε_t^{λ} is given by:

$$p_{\lambda}(s,t) = \sum_{s=s_j \in S_t^{\lambda}} \eta(s_j) e^{-|s_j - \varepsilon_{t-1}^{\lambda}|}, \quad (1)$$

where $\eta(s_j) = e^j / \sum_{k=1}^{w} e^k$ is the weight of $s_j \in S_t^{\lambda}$ (the most recent hidden state has the most significant influence in predicting the next state). The estimated state for the current step is the state with maximum likelihood:

$$\varepsilon_t^{\lambda} = \underset{s \in S_t^{\lambda}}{\operatorname{argmax}}\, p_{\lambda}(s,t) \quad (2)$$

3.2 Agent Processing Load Model

According to schema theory [11], multiple elements of information can be chunked as single elements in cognitive schemas. A schema can hold a huge amount of information, yet is processed as a single unit. We adapt this idea and assume that agent i's estimation of agent j's processing load at time step t is a function of two factors: the number of chunks $c_j(t)$ and the total number $s_j(t)$ of information being considered by agent j. If $c_j(t)$ and $s_j(t)$ are observable to agent i, agent i can employ a Windowed-HMM approach as described in Section 3.1 to model and estimate agent j's instantaneous processing load.

In the study reported below, we also used 5-state HMM models for agent processing load. With the 5 hidden states similar to the HMM models adopted for human cognitive load, we employed multivariate Gaussian observation probability distributions for the hidden states.

3.3 HAP's Processing Load Model

As discussed above, a Human-Agent-Pair (HAP) is viewed as a unit when teaming up with other HAPs. The processing load of a HAP can thus be modeled as the co-effect of the processing load of the agent and the cognitive load of the human partner as captured by the agent.

Suppose agent A_i has models for its processing load and its human partner H_i's cognitive load. Denote the agent processing load and human cognitive load of HAP $\langle H_i, A_i \rangle$ at time step t by μ_t^i and ν_t^i, respectively. Agent A_i can use μ_t^i and ν_t^i to estimate the load of $\langle H_i, A_i \rangle$ as a whole. Similarly, if μ_t^j and ν_t^j are observable to agent A_i, it can estimate the load of $\langle H_j, A_j \rangle$. For model simplicity, we still used 5-state HMM models for HAP processing load, with the estimated hidden states of the corresponding agent processing load and human cognitive load as the input observation vectors.

Building a load estimation model is the means. The goal is to use the model to enhance information sharing performance so that a team can form better shared mental models (e.g., to develop inter-agent role expectations in their collaboration), which is the key to high team performance.

3.4 Load-Sensitive Information Processing

Each agent has to adopt a certain strategy to process the incoming information. As far as resource-bounded agents are concerned, it is of no use for an agent to share information with teammates who are already overloaded; they simply do not have the capacity to process the information.

Consider the incoming information processing strategy as shown in Table 1. Agent A_i (of HAP_i) ignores all the incoming information when it is overloaded, and processes all the incoming information when it is negligibly loaded. When it is heavily loaded, A_i randomly processes half of the messages from those agents which are the first $1/q$ teammates appeared in its communication queue; when it is fairly loaded, A_i randomly processes half of the messages from any teammates; when it is slightly loaded, A_i processes all the messages from those agents which are the first $1/q$ teammates appeared in its communication queue, and randomly processes half of the messages from other teammates.

Table 1: Incoming information processing strategy

HAP_i Load	*Strategy*
Overly	Ignore all shared info
Heavily	Consider every teammate $A \in [1, \lceil \frac{1}{q}\|Q\| \rceil]$, randomly process half amount of info from A; Ignore info from any teammate $B \in (\frac{1}{q}\|Q\|, \|Q\|]$
Fairly	Process half of shared info from any teammate
Slightly	Process all info from any $A \in [1, \lceil \frac{1}{q}\|Q\| \rceil]$; For any teammate $B \in (\frac{1}{q}\|Q\|, \|Q\|]$ randomly process half amount of info from B
Negligibly	Process all shared info
$\not\Rightarrow HAP_j$	Process all info from HAP_j if it is overloaded

*Q is a FIFO queue of agents from whom this HAP has received information at the current step; q is a constant known to all.

To further encourage sharing information at the right time, the last row of Table 1 says that HAP_i, if having not sent information to HAP_j who is currently overloaded, will process all the information from HAP_j. This can be justified from resource allocation perspective: an agent can reallocate its computing resource reserved for communication to enhance its capacity of processing information. This strategy favors "never sending information to an overloaded teammate", and it suggests that estimating and exploiting others' loads can be critical to enable an agent to share the right information with the right party at the right time.

4. SYSTEM IMPLEMENTATION

SMMall (Shared Mental Models for all) is a cognitive agent architecture developed for supporting human-centric collaborative computing. It stresses human's role in team activities by means of novel collaborative concepts and multiple representations of context woven through all aspects of team work. Here we describe two components pertinent to the experiment reported in Section 5: multi-party communication and shared mental maps (a complete description of the SMMall system is beyond the scope of this paper).

4.1 Multi-Party Communication

Multi-party communication refers to conversations involving more than two parties. Aside from the speaker, the listeners involved in a conversation can be classified into various roles such as addressees (the direct listeners), auditors (the intended listeners), overhearers (the unintended but anticipated listeners), and eavesdroppers (the unanticipated listeners). Multi-party communication is one of the characteristics of human teams. SMMall agents, which can form Human-Agent-Pairs with human partners, support multi-party communication with the following features.

1. SMMall supports a collection of multi-party performatives such as MInform (multi-party inform), MAnnounce (multi-party announce), and MAsk (multi-party ask). The listeners of a multi-party performative can be addressees, auditors, and overhearers, which correspond to 'to', 'cc', and 'bcc' in e-mail terms, respectively.

2. SMMall supports channelled-communication. There

are three built-in channels: agentTalk channel (inter-agent activity-specific communication), control channel (meta communication for team coordination), and world channel (communication with the external world). An agent can fully "tune" to a channel to collect messages sent (or cc, bcc) to it. An agent can also *partially* tune to a channel to get statistic information about the messages communicated over the channel. This is particularly useful if an agent wants to know the communication load imposed on a teammate.

4.2 Shared Belief Map & Load Display

A concept "shared belief map" has been proposed and implemented into SMMall; this responds to the need to seek innovative perspectives or concepts that allow group members to effectively represent and reason about shared mental models at different levels of abstraction. As described in Section 5, humans and agents interacted through shared belief maps in the evaluation of HMM-based load models.

A shared belief map is a table with color-coded *info-cells*—cells associated with information. Each row captures the belief model of one team member, and each column corresponds to a specific information type (all columns together define the boundary of the information space being considered). Thus, info-cell C_{ij} of a map encodes all the beliefs (instances) of information type j held by agent i. Color coding applies to each info-cell to indicate the number of information instances held by the corresponding agent.

The concept of shared belief map helps maintain and present a human partner with a synergy view of the shared mental models evolving within a team. Briefly, SMMall has implemented the concept with the following features:

1. A context menu can be popped up for each info-cell to view and share the associated information instances. It allows selective (selected subset) or holistic info-sharing.

2. Mixed-initiative info-sharing: both agents and human partners can initiate a multi-party conversation. It also allows third-party info-sharing, say, A shares the information held by B with C.

3. Information types that are semantically related (e.g., by inference rules) can be closely organized. Hence, nearby info-cells can form meaningful plateaus (or contour lines) of similar colors. Colored plateaus indicate those sections of a shared mental model that bear high overlapping degrees.

4. The perceptible color (hue) difference manifested from a shared belief map indicates the information difference among team members, and hence visually represents the potential information needs of each team member (See Figure 3).

SMMall has also implemented the HMM-based models (Section 3) to allow an agent to estimate its human partner's and other team members' cognitive/processing loads.

As shown in Fig. 3, below the shared belief map there is a load display for each team member. There are 2 curves in a display: the blue (dark) one plots human's instantaneous cognitive loads and the red one plots the processing loads of a HAP as a whole. If there are n team members, each agent needs to maintain $2n$ HMM-based models to support the load displays. The human partner of a HAP can adjust her cognitive load at runtime, as well as monitor another HAP's agent processing load and its probability of processing the information she sends at the current time step. Thus, the more closely a HAP can estimate the actual processing loads of other HAPs, the better information sharing performance the HAP can achieve.

Figure 3: Shared Mental Map Display

In sum, shared belief maps allow the inference of who needs what, and load displays allow the judgment of when to share information. Together they allow us to investigate the impact of *sharing the right info. with the right party at the right time* on the evolving of shared mental models.

4.3 Metrics for Shared Mental Models

We here describe how we measure team performance in our experiment. We use "mental model overlapping percentage (MMOP)" as the base to measure shared mental models. MMOP of a group is defined as the intersection of all the individual mental states relative to the union of individual mental states of the group. Formally, given a group of k agents $G = \{A_i | 1 \leq i \leq k\}$, let $B_i = \{I_{im} | 1 \leq m \leq n\}$ be the beliefs (information) held by agent A_i, where each I_{im} is a set of information of the same type, and n (the size of information space) is fixed for the agents in G, then

$$MMOP(G) = \frac{100}{n} \sum_{1 \leq m \leq n} \left(\frac{|\cap_{1 \leq i \leq k} I_{im}|}{|\cup_{1 \leq i \leq k} I_{im}|} \right). \quad (3)$$

First, a shared mental model can be measured in terms of the "distance" of averaged subgroup MMOPs to the MMOP of the whole group. Without losing generality, we define paired SMM distance (subgroups of size 2) $D_2^\top$ as:

$$D_2^\top(G) = \sum_{1 \leq i < j \leq k} (MMOP(\{A_i, A_j\}) - MMOP(G))^2. \quad (4)$$

The MMOP of a subgroup is always larger than the MMOP of the whole group. Intuitively, the larger distance from the MMOP of a subgroup to that of the whole group, the more overlapping mental models the subgroup shares. This notion can be used to measure the tightness of an emerging subgroup or guide the process of team coalition.

Second, due to communication or information processing limits, each individual's subjective measure of the group's MMOP can be very different from the group's MMOP measured objectively from external. A shared mental model can thus be measured in terms of the "closeness" of individuals' measure of the group's MMOP to the objective measure. Let *MMOP(G)* and $MMOP_i(G)$ be the objective measure and agent A_i's subjective measure of the group's shared mental models, respectively. We define SMM deviation $D_\perp$ as:

$$D_\perp(G) = \sum_{1 \leq i \leq k} (MMOP_i(G) - MMOP(G))^2. \quad (5)$$

Obviously, $D_\perp$ measures the coherency of the whole group: the smaller the better.

Third, shared mental models evolve over time. A shared mental model can be measured in terms of the "stableness" of the instantaneous measures of MMOP, $D^\top$, or $D_\perp$

over time. High performing teams can often maintain their shared mental models such that the MMOP is stable at an acceptable level as activities proceed, while the MMOP measure of the shared mental models of a low-performing teams can fluctuate or decrease notably over time.

5. EXPERIMENT EVALUATION

In this section we describe the experiments conducted to evaluate the load estimation models and the shared belief map concept for developing team shared mental models.

5.1 Problem Description

The members of a HAP team (i.e., a team composed of Human-Agent-Pairs) are situated in a dynamic environment. Due to their different (maybe overlapping) observability, at each time step they may get different situational information. The goal of a HAP team is to selectively share information in a timely manner to develop global situation awareness (say, for making critical decisions).

Each run of the experiment has 45 time steps; each time step lasts 15 seconds. A time step starts with certain info-cells of the shared belief map being flashed quickly (for 2 s). The flashed cells are exactly those with newly available information that should be shared at that time step. An info-cell is frozen at the end of a time step: the associated information is no longer sharable. This requires that the newly available information be shared in time, not later.

The human and agent of a HAP assume different roles. An agent governs group communication and processes messages to update the shared belief map on its display. An agent, with a pre-trained HMM-based cognitive load model for its human partner and a processing load model for each of the other HAPs in the team, also estimates and displays their instantaneous loads. Human subjects need to perform a primary task and a secondary task. The *secondary task* of a human subject is to remember and mark the cells being flashed (not necessarily in the exact order) by left mouse clicks. Secondary task performance at step t is thus measured as the number of cells marked (remembered) correctly at t, which is taken as the observable state of the HMM-based cognitive load model of that human subject. The *primary task* of a human subject is to *build a shared mental model* of the dynamic situation by sharing the right information with the right party at the right time. To share the information associated with an info-cell, a human subject needs to click the right mouse button on the cell, and select the receiving teammate(s) from the popup menu.

There are costs associated with information sharing. Communications among HAPs is done by the corresponding SMMall agents, which have both limited capacity n_{in} for processing incoming information and limited outgoing communication capacity n_{out}. Thus, depending on the current HAP load, an agent may randomly ignore part or all of the incoming information (having no effect on the establishment of shared mental models). On the other hand, each time step at most n_{out} number of information-sharing commands can be effective; more than that contribute nothing to the establishment of shared mental models. Sending information to an overloaded teammate will waste the capacity that otherwise can be used to share information with a less loaded teammate. This means that at each time step a human subject has to carefully go through three cognitive decisions: whether the information under consideration needs to be shared (i.e., whether it is associated with an info-cell just flashed), whether a team member is the right party to share the information with (i.e., whether it really needs the information), and whether this is the right time to share (i.e., whether the team member is currently overloaded).

The above description applies to HAP teams. For teams composed of SMMall agents only, the agents will take all the roles played by an agent or a human partner in HAP teams.

5.2 Experiment Design

To investigate the impacts of the HMM-based load models on the evolution of shared mental models (SMM), we conducted experiments for both Agent teams and HAP teams, where agent teams involve agent processing load models only, HAP teams involve models of HAP processing load (i.e., the co-effect of agent processing load and human cognitive load). To get insights on how load predictions and multi-party communication may affect the performance of forming SMM, we designed 3 Agent teams (TA1, TA2, TA3) and 3 HAP teams (TH1, TH2, TH3), where all agents adopt the strategy in Table 1 to send and process information. When sharing information, teams of type 1 (TA1, TH1) ignore load predictions; teams of type 2 (TA2, TH2) consider load predictions; teams of type 3 (TA3, TH3) follow load predictions more strictly in the sense that the agents will further group the receivers of a multi-party message (MInform) by their loads and split the message into multiple messages with their receivers having the same load. For example, given that agents A_1, A_2, A_3 have load 1, 2, 1, respectively. An agent A_0 in a team of TA2 may send one multi-party message, while an agent A_0' in a team of TA3 will send two messages (one to A_2, one to A_1 and A_3). In addition, we controlled agents' outgoing communication capacities by varying from 6, 8, to 10.

Due to constraints on communication capacity and processing capacity, an agent can be inaccurate when tracking other teammates' mental models. In order to measure the *actual* shared mental models, a special SMMall agent named 'OmniAll' was added to each team to monitor inter-agent communications and record the actual effects of information sharing on each agent's mental model. This realizes a way, as suggested by Klimoski [6], to measuring the degree of overlap in immediate, intermediate, and long-term situation awareness zones held by group members.

We also recorded instantaneous information sharing utility, which is defined as follows. At each time step, let $T = \langle T_0, T_1, T_2, T_3, T_4 \rangle$ be a sequence of sets, where T_0, T_1, T_2, T_3, and T_4 are sets of teammates whose current load states are "negligibly", "slightly", "fairly", "heavily", and "overly" respectively. Let S be the set of information-sharing commands issued by a human partner at the current step. Let $M_i = \{T_k \in T | k \leq i, T_k \neq \emptyset\}$. Instantaneous info-sharing utility is defined as $\sum_{c \in S} s\_value(c)/|S|$, where

$$s\_value(c) = \begin{cases} 0 & receiver(c) \in T_4 \\ 0 & c \text{ is known to } receiver(c) \\ 1/|M_i| & receiver(c) \in T_i, i \neq 4 \end{cases} \quad (6)$$

In sum, this study involved 18 types of teams, each team had 4 members, and each team type was tested by 10 domain scenarios. 30 human subjects were recruited for HAP teams. The experiment results are plotted in Figures 4, 6, and 7. We next present our findings in this study.

5.3 Load Estimation Betters Info-sharing

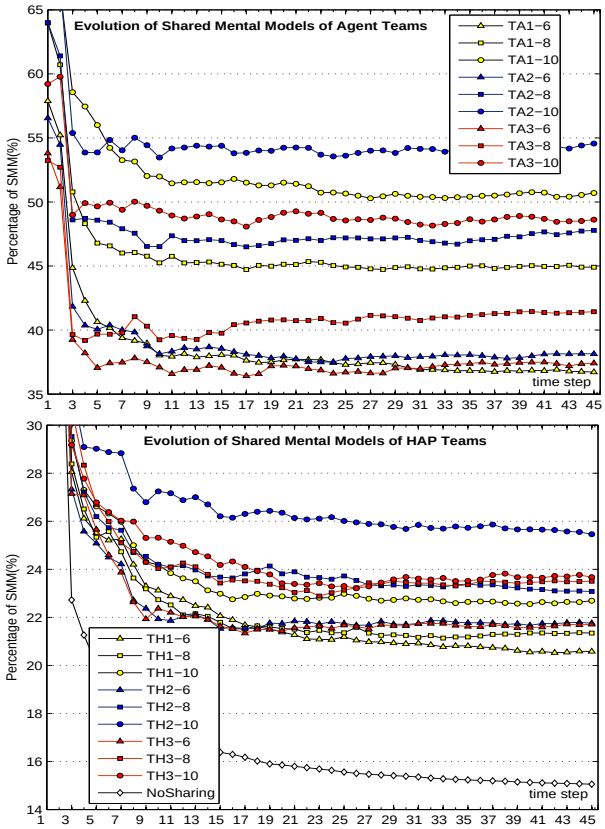


Figure 4: Evolution of SMMs.

Consider teams of type 1 (TA1, TH1) and teams of type 2 (TA2, TH2). First look at the performance of HAP teams in Fig. 4, we have: (1) For each team type, the performance (percentage of SMM overtime) averaged over 10 teams increased as communication capacity increased (TH1-6<TH1-8<TH1-10, TH2-6<TH2-8<TH2-10). (2) The averaged performance of TH2 teams performed consistently better than the TH1 teams for each capacity setting (TH2-6>TH1-6, TH2-8>TH1-8, TH2-10>TH1-10), and the performance difference of TH1 and TH2 teams increased as communication capacity increased. This indicates that, other things being equal, the benefit of exploiting load estimation when sharing information becomes more significant when communication capacity is larger. From Fig. 4 the same findings can be derived for the performance of agent teams.

In addition, the results also show that the SMMs of each team type were maintained steadily at a certain level after about 20 time steps. However, to maintain a SMM steadily at a certain level is a non-trivial team task. The performance of teams who did not share any information (the 'NoSharing' curve in Fig. 4) decreased constantly as time proceeded.

5.4 Multi-Party Communication for SMM

We now compare teams of type 2 and type 3 (which splits multi-party messages by receivers' loads). As plotted in Fig. 4, for HAP teams, the performance of team type 2 for each fixed communication capacity was consistently better than team type 3 (TH3-6≤TH2-6, TH3-8<TH2-8, TH3-10<TH2-10); the difference became more significant as communication capacity increased. These also hold for the Agent teams (upper one in Fig. 4). Actually, the performance of type 3 agent teams was even worse (for each fixed capacity) than the performance of type 1 agent teams.

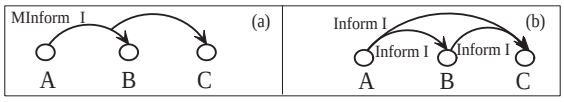


Figure 5: Multi-party messages.

This can be explained by the difference of two-party and multi-party communications. In SMMall, in order to enable team-level inference, each agent maintains an internal model of every team member's mental model (beliefs). According to the semantics of MInform (multi-party inform), when A MInforms I to others, assuming all the receivers and overhearers will accept I, A will update its internal model of their beliefs; each of the receivers (overhearers), upon getting the message, will update its own beliefs as well as its model of the sender's and all the other receivers' beliefs.

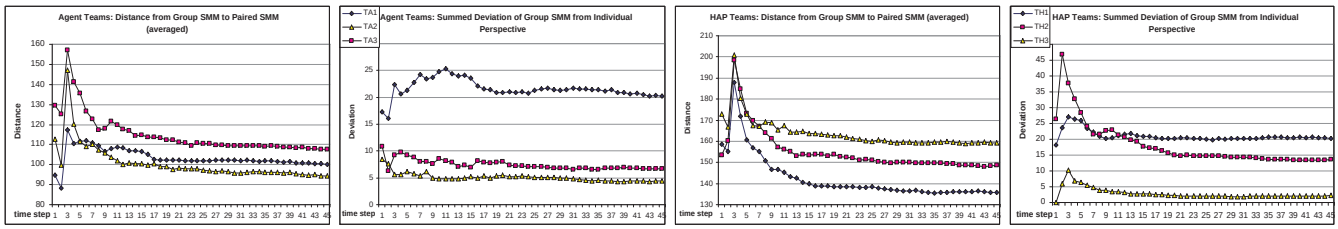
Compared to two-party performatives, multi-party performatives are preferable for forming shared mental models. As illustrated in Fig. 5(a), agent A only needs to perform MInform once (with B and C being receivers) to achieve the common knowledge of the shared belief about I (It consumes 2 of A's communication resources, one for each receiver). However, to achieve the same effects using Inform (Fig. 5(b)), it is hard to form team-level SMM especially when the team size is big (missing one Inform will nullify all others' efforts). Moreover, although agent A still consumes 2, the whole team needs more resources (3 in this case).

However, splitting multi-party messages by receivers' loads does enhance subgroup SMMs. In Fig. 6 we plotted the other two measures of SMMs (distance and closeness as defined in Sec. 4.3). For HAP teams, TH3>TH2>TH1 holds in Fig. 6(c) (larger distances indicate better subgroup SMMs), and TH3<TH2<TH1 holds in Fig. 6(d) (smaller deviations indicate higher coherency of the whole team). Thus, HAP teams of type 3 achieved better subgroup SMMs, and their team members had higher coherent view of group SMMs than teams of other types. For agent teams, TH3>TH1>TH2 holds in Fig. 6(a), and TH2<TH3<TH1 holds in Fig. 6(b). Thus, agent teams of type 3 achieved better subgroup SMMs, and their team members had much higher coherent view of group SMMs than teams of type 1 (although slightly worse than team type 2).

Hence, generally, multi-party communication encourages the forming/evolving of team SMMs. When a group of agents can be partitioned into subteams, splitting messages by their loads can be more effective for subteam SMMs.

5.5 The HMM-based Cognitive Model

To validate the HMM-based cognitive load model is extremely difficult because detecting the real, noise-resistant cognitive load of human beings is far beyond the current technology. As an indirect judgment, we plotted the regression fitted lines for the means of information sharing utility of HAP teams with and without load displays. For the HAP teams with load displays there is a strong negative linear relationship between info-sharing utility and cognitive load levels (Pearson correlation coefficient is $-\sqrt{0.899} = -0.948$ with P-Value $= 0.014$). Because the info-sharing utility measure and the cognitive load measure are indicators of primary task performance and secondary task performance, respectively, their linear relationship complies with cognitive studies that secondary task performance can be used



(a) (b) (c) (d)

Figure 6: The distance of subgroup SMMs and the closeness of individual views of team SMMs.

to explain primary task performance. However, for teams without load displays there is a strong quadratic rather than linear relationship (correlation coefficient is $\sqrt{0.974} = 0.987$ with P-Value = 0.015). This indicates that the information sharing performance of a HAP team can be significantly affected by both the human subject's own cognitive load, and the awareness of other team members' load. Knowing of others' load (by estimation) will reduce the quadratic relation to a linear relation.

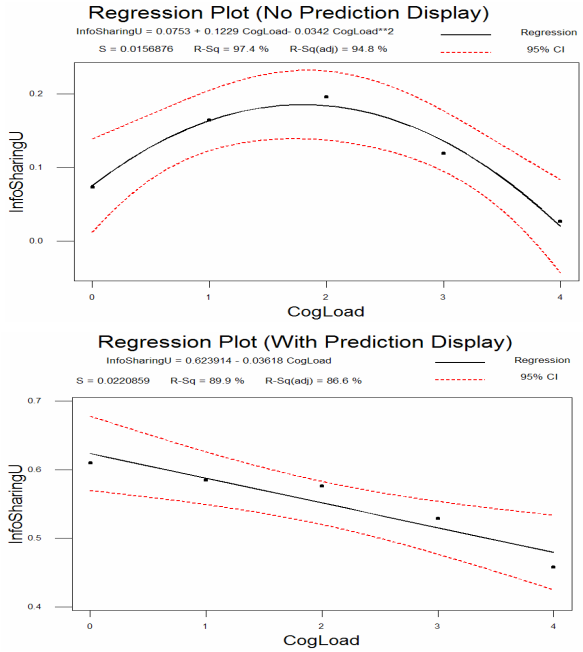


Figure 7: Info-sharing utility by cognitive loads.

6. CONCLUSION

Recent research attention on human-centered teamwork highly demands the design of agent systems as cognitive aids that can model and exploit human partners' cognitive capacities to offer help unintrusively. In this paper, we investigated several factors surrounding the challenging problem of evolving shared mental models of teams composed of human-agent-pairs. The major contribution of this research includes (1) HMM-based load models were proposed for an agent to estimate its human partner's cognitive load and other HAP teammates' processing loads; (2) The shared belief map concept was introduced and implemented. It allows group members to effectively represent and reason about shared mental models; (3) Experiments were conducted to evaluate the HMM-based cognitive/processing load models and the impacts of multi-party communication on the evolving of team SMMs. The usefulness of shared belief maps was also demonstrated during the experiments.

7. REFERENCES

[1] A. D. Baddeley. Working memory. *Science*, 255:556–559, 1992.

[2] J. Bradshaw, M. Sierhuis, A. Acquisti, Y. Gawdiak, D. Prescott, R. Jeffers, N. Suri, and R. van Hoof. What we can learn about human-agent teamwork from practice. In *Workshop on Teamwork and Coalition Formation at AAMAS'02)*, Bologna, Italy, 2002.

[3] P. R. Cohen and H. J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.

[4] X. Fan, B. Sun, S. Sun, M. McNeese, and J. Yen. RPD-enabled agents teaming with humans for multi-context decision making. In *Proceedings of AAMAS'06*, pages 34–41. ACM Press, 2006.

[5] B. Grosz and S. Kraus. Collaborative plans for complex group actions. *Artificial Intelligence*, 86:269–358, 1996.

[6] R. Klimoski and S. Mohammed. Team mental model: Construct or metaphor? *Journal of Management*, 20(2):403–437, 1994.

[7] A. Lang. The limited capacity model of mediated message processing. *J. of Comm.*, Winter:46–70, 2000.

[8] R. G. Lord and K. J. Maher. Alternative information processing models and their implications for theory, research, and practice. *The Academy of Management Review*, 15(1):9–28, 1990.

[9] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[10] E. Norling. Folk psychology for human modelling: Extending the BDI paradigm. In *Proceedings of AAMAS'04*, pages 202–209, 2004.

[11] F. Paas and J. V. Merrienboer. The efficiency of instructional conditions: an approach to combine mental-effort and performance measures. *Human Factors*, 35:737–743, 1993.

[12] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. V. Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1):63–71, 2003.

[13] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[14] W. Rouse, J. Cannon-Bowers, and E. Salas. The role of mental models in team performance in complex systems. *IEEE Trans. on Sys., man, and Cyber*, 22(6):1296–1308, 1992.

[15] J. Sweller. Cognitive load during problem solving: effects on learning. *Cog. Science*, 12:257–285, 1988.

[16] B. Xie and G. Salvendy. Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics*, 4:213–242, 2000.