

Personal Assistants for Human Organizations

Steven Okamoto
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
USA
Phone: +1 412-268-8780
sokamoto@cs.cmu.edu

Katia Sycara
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
USA
Phone: +1 412-268-8825
katia@cs.cmu.edu

Paul Scerri
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
USA
Phone : +1 412-268-2145
pscerri@cs.cmu.edu

Personal Assistants for Human Organizations

Steven Okamoto

Carnegie Mellon University, USA

Katia Sycara

Carnegie Mellon University, USA

Paul Scerri

Carnegie Mellon University, USA

ABSTRACT

Intelligent software personal assistants are an active research area with the potential to revolutionize the way that human organizations operate, but there has been little research quantifying how they will impact organizational performance or how organizations will or should adapt in response. In this chapter we develop a computational model of the organization to evaluate the impact different proposed assistant abilities have on the behavior and performance of the organization. By varying the organizational structures under consideration, we can identify which abilities are most beneficial, as well as explore how organizations may adapt to best leverage the new technology. The results indicate that the most beneficial abilities for hierarchical organizations are those that improve load balancing through task allocation and failure recovery, while for horizontal organizations the most beneficial abilities are those that improve communication. The results also suggest that software personal assistant technology will facilitate more horizontal organizations.

INTRODUCTION

Intelligent software personal assistants (SPAs) are one of the most exciting applications for organizational multi-agent systems. A **software personal assistant (SPA)** is an agent that acts to support a user in a human organization by automating individual tasks and facilitating coordination with other members of the organization. Recent and current research has looked at developing SPAs for a diverse range of domains, including emergency response and military teams, office environments, factory floors, and even outer space. The envisioned SPAs possess a wide range of abilities, such as scheduling joint activities (Dent et al, 1992; Garrido & Sycara, 1996; Modi et al, 2004), sharing key information (Wagner et al, 2004), monitoring and reminding individuals of key timepoints (Chalupsky et al, 2001), filtering incoming communication (Maes, 1994),

assisting in negotiation decision support (Li et al., 2006), and even ordering lunch (Chalupsky et al, 2001).

Software personal assistants stand to benefit from multi-agent organization research in two ways. First, as large-scale, complex multi-agent systems, SPA deployments are natural candidates for organization-centric engineering approaches to manage control and coordination complexity. As SPA-enabled organizations become more commonplace, the need for an organization-centric approach will only become more apparent, because SPA interactions between different organizations will require that these systems operate flexibly, robustly, and securely as open systems. Secondly, no matter what engineering approach is chosen for an SPA system, the SPAs will be situated in human organizations with specific organizational constraints, and thus the SPAs must be able to represent and reason about those constraints in order to operate successfully and transparently. The representation and reasoning of organizational structures and norms is currently one of the hottest areas of agent organization research (Grossi et al., 2007; Vasconcelos et al., 2007).

While significant technical challenges remain in developing SPAs, many of these issues are the subject of recent and current research and will not be discussed extensively in this chapter. Instead, we will focus on the crucial issue of how human organizations will be affected by the use of this technology, which has gone largely unexamined. This represents a significant gap in the current research, as it is likely that SPAs will have a revolutionary effect on the way human organizations operate, just as previous information technology innovations such as personal computers, corporate databases, and e-mail revolutionized the way organizations operate. In addition, the issue of which of the many envisioned SPA abilities are most useful for improving the efficiency and effectiveness of an organization is only poorly understood. This is a possibly costly oversight, as history is replete with examples of technological innovations, including early SPA systems, that have unexpected and even undesirable impacts when coupled with existing organizational practices and behaviors.

Our goals for this chapter are three-fold: first, to develop a conceptual framework that can be used to quantifiably evaluate proposed SPA technologies; second, to quantify the impacts proposed SPA abilities will have on existing organizations, in order to provide input to SPA designers on which abilities are most promising to pursue; and third, to explore how organizations may best be redesigned to leverage the SPA technology, in order to provide input to SPA adopters on how to best apply the technology. The approach taken in this chapter lies at the crossroads of agent organization modeling and computational organization theory. We develop a **computational model** of the organization. This methodology of computational modeling is similar to that used for quantifying performance in organizational theory (Carley 1994; Prietula & Carley, 1998). We evaluate the impact different proposed SPA abilities have on the behavior and performance of the organization. Because SPAs frequently affect detailed work activities, and their broader influence on the organization is unclear, we must model the organization at a fine-grained level of detail, capturing, for example, communication paths, decision making, etc., in order to see their effects. To that end, we have created an abstract simulation environment that takes proposed organizational models and tasks to be performed by the organization and computes key properties of task execution, including how well and how quickly the organization performed the task and how

robustly it handled individual failures. The simulation captures important aspects of the operation of the organization, such as non-determinism and cognitive limits of individual members of the organization, but abstracts away domain level details, making it feasible to evaluate many instances of organizations.

One of the major difficulties in evaluating SPA abilities is that they are still an area of ongoing research. Because one of our goals is to provide input to SPA designers before the SPA capabilities have been fully developed, it is not possible to directly model the specific mechanisms by which SPAs will operate. Instead, we abstract away the details of the SPA mechanisms and instead use the effects of those mechanisms on the behavior of individual humans to determine the impact on the organization. For example, instead of directly modeling the specific ways in which SPAs will increase the rate at which humans can make effective decisions, we instead model the SPAs' impact as increasing the decision making rate within the organization, with the understanding that any SPA mechanism that has the same effect will have similar effects on organizational performance. This allows us to evaluate the many proposed SPA abilities without having to solve the hard and open problems required to implement those abilities for different organizational contexts. By referencing ongoing projects and previously published literature, we have identified a set of key abilities that are being developed, including information sharing, task allocation, automated monitoring and supervision, communication management, joint activity coordination, decision support, and recovery from unexpected failures. Using our computational model of the organization, we evaluate the impacts of different SPA abilities individually and in combination.

When embracing a new technology, organizations will first deploy it in lieu of existing technologies or practices, then gradually adapt to better utilize the new technology. This can have a transformative effect on the way organizations are structured and operate. For this reason, it is not sufficient to merely study the impact SPAs will have on existing organizations, but also to see how organizations might change and adapt in the presence of SPAs to best leverage the technology. In order to investigate this, we compare the designs of organizations before and after SPA deployment to see which structures and SPA abilities are most beneficial. Our results indicate that the most beneficial SPA abilities for horizontal, decentralized organizations are those that facilitate improved communication, while for hierarchical, centralized organizations, the most beneficial SPA abilities are those that improve load balancing through task allocation and failure recovery.

BACKGROUND

Research on SPAs has generally focused on developing SPAs for a wide range of environments, including office settings (Chen & Sycara, 1998; Modi et al, 2005; Payne et al, 2002), disaster response domains (Schurr et al, 2005), and military operations (Lenox et al, 2000b; Wagner et al, 2004), rather than on studying their impact on human organizations. This research into developing SPAs can be roughly partitioned into three categories. The first category focuses on technical issues underlying specific abilities, using techniques such as Markov decision processes (Varakantham et al., 2005), constraint satisfaction and constraint optimization (Modi, 2004; Maheswaran et al., 2004), and decision trees (Mitchell, 1994). Performance evaluations in these approaches

tend to focus solely on the performance of the SPA (measured by computation time, learning accuracy, etc.), and completely ignore the human user or the organization.

The second category considers the interaction between an SPA and its user. This research focuses on developing and evaluating different user interfaces for SPAs, taking into account factors such as (McCrickard & Chewar, 2003; Horvitz, 1999). This usually involves human experiments interacting in which a user interacts with a prototype SPA. While these studies consider the cognitive capabilities and limitations of a human user and how these can be affected by an SPA, they consider only the *individual* performance of a single user and its SPAs, instead of the organizational performance of an SPA-enabled organization. However, results from these studies are complementary to the approach taken in this chapter, as they quantify the effects SPAs have on individual performance, while we begin with the impacts SPAs have on individual performance and quantify the effect this has on the collective performance of the organization as a whole.

The third category includes systems of agents that have actually been built and tested in assisting humans (Chalupsky et al, 2001; Lenox et al, 2000a; Sycara & Lewis, 2004), and is a far smaller category than the other two. These studies also involve experiments with human subjects, but several subjects interact with the system at a time, assisted by their respective SPAs. This is closest in spirit to our objective, however, these experiments are generally very limited in size and scope, and frequently are not conducted on the true end-users of the technology. This is natural because of the impracticality of deploying a full SPA system in a large organization. In the largest and longest-running SPA experiment to date, an SPA system (the “Electric Elves”) was built and deployed in an office environment and used for several months (Chalupsky et al., 2001). However, even this system was operated with fewer than 10 users. While experiments such as these are important for the eventual development and deployment of SPAs in the workplace, they are impractical for large scale tests and are inherently restricted to testing existing SPA technology.

Computational organization theory has focused specifically on the human and organizational performance factors that SPA-oriented research has largely ignored. This research area uses **computational models** of human organizations to answer questions of organizational behavior and performance (Carley & Prietula, 1994; Prietula et al., 1998). It is widely accepted as a practical way to conduct experiments on large scale human organizations, especially using “what-if” analysis to consider the effects hypothetical changes will have on organizational behavior (Carley, 1995). This approach has been used to evaluate the effects of different forms of information technology on specific organizations (Levitt et al. 1994; Majchrzak & Finley, 1995), but these approaches require a detailed model of how the information technology works, which is not practical in the case of yet undeveloped SPAs.

One aspect of SPA design that has been neglected is the explicit representation and reasoning about organizational concepts. For example, while the Electric Elves project was generally considered a success, the SPAs were observed to have several serious shortcomings, because they had not been designed with an organizational awareness (Tambe et al., 2006). Instead, machine learning methods were used to generalize rules to guide their operation, and in many cases the learned rules violated the norms of the human organization in which they were situated. The system designers included ad hoc rules in an attempt to repair this shortcoming, but were only partially successful.

Although the approach taken in this chapter is agnostic as to the specific engineering methodology used for SPAs, we recognize that an organization-centric approach to agent control could solve many of the issues that plagued the Electric Elves project by providing the SPAs with the ability to reason about the organizational norms in choosing its actions. Even if an agent-centric design process of the SPAs system as a whole was used, SPAs could still greatly benefit from the formal organizational languages (Grossi et al. 2007; Serrano & Saugar, 2007) and organizational frameworks (Dignum et al., 2004) in order to adhere to the norms of the human organizations in which they are situated. For a more detailed exposition on how SPAs may be made organizationally-aware, please see chapter 22 “Autonomous agents adopting organization rules,” by Bob van der Vecht and Frank Dignum.

The approach presented in this chapter uses organizational models in a different way than they are conventionally used in multi-agent systems. The conventional usage (which is also reflected in most of the other chapters of this book) is to use organizational models to provide organization-centric control of a distributed agent system, while remaining agnostic about the evaluation methodology. In contrast, an organizational model is used in this chapter to simulate the operation of a human organization in order to evaluate the organizational performance, while remaining agnostic about the control methodology. It is important to note that these two approaches may be used complementarily; the approach presented in this chapter is certainly capable of evaluating organization-centric SPA systems, but it is also capable of evaluating agent-centric SPA systems, which is a key advantage, as most SPA systems to date have been designed from an agent-centric perspective.

Other recent research efforts have focused on evaluating the organizational performance of multi-agent systems (Grossi et al., 2006; Horling, 2006). An organizational design language expressing quantitative relationships that exist between organizational elements was provided in (Horling, 2006), along with a framework for quantifying various organizational performance characteristics. In (Grossi et al., 2006), aspects of organizational performance are determined by examining graph theoretic properties of the structure of the organizational model. This allows several general measures of performance to be made, but relies on an explicit organization-centric model of control and the performance measures are static with respect to the organizational structure. While these works present important advances in moving from agent-centric to organization-centric performance evaluation, they are both focused on pure agent organizations, instead of specifically addressing hybrid organizations such as an SPA-enabled organization. Preliminary work on understanding the impact of software personal assistants was described in (Okamoto et al., 2006).

The adoption of information technology by organizations generally follows three phases (Carley, 1996). In the first phase, the new technology is used as a simple substitute for existing technologies and practices. In the second phase, the organization members adjust their behavior as they become more familiar with the properties of the new technology. Finally, in the third phase, the **organizational structure** is changed to better fit the new behavior. Three notable and well-studied technologies that have adhered to this general progression are electronic marketplaces (Malone, 1987), e-mail (Yates & Orlikowski, 1992; Garton & Wellman, 1993), and corporate databases (Carley, 2002).

Electronic marketplaces initially replaced existing intra-organization coordination procedures between different divisions, thereby reducing costs (Malone, 1987). As these savings were realized, usage of electronic marketplaces increased. Finally, the coordination costs were sufficiently reduced that many organizations chose to outsource the work previously done in-house, and use electronic marketplaces for inter-organizational coordination.

Similarly, e-mail was initially used as a replacement for telephone calls and written memoranda (Yates & Orlikowski, 1992). This combined the asynchronous and permanent nature of memoranda with the speed and interactivity of telephone calls. This led changes in organizational behavior, including status equalization and often an increased volume of communication through e-mail as organizations moved through the second phase of adoption (Garton & Wellman, 1993). Finally, in the third phase organizations were restructured to take advantage of the new communication potential of e-mail, with less centralized, less hierarchical structures (Garton & Wellman, 1993; Ahuja & Carley, 1998).

It is generally assumed that information technology reduces coordination costs and improves organizational performance, an assumption that is supported by the studies of electronic marketplaces and e-mail. However, it has also long been observed that information technology may have a negative effect on organizational performance (Ackoff, 1967). A study of corporate databases provides a more recent cautionary warning (Carley, 2002). In this case a database was deployed to increase the retention and accessibility of organizational knowledge that had previously been retained only by a handful of experts in the organization. The initial phase of adoption led to a shift in reliance from these experts to the database system. However, in the second phase, people stopped updating the database, assuming that someone else would, leading to an eventual loss of organizational knowledge. This highlights the point that it is essential to account for the interactions between a proposed technology and the organization's dynamics in order to avoid deleterious effects.

APPROACH

In this section we elaborate on the general approach to evaluating the impact of SPAs on human organizations. We will also provide a concrete example for which we will later provide results. Because direct experimentation with real SPA-enabled organizations is impractical, we construct computational models of human organizations to evaluate organizational performance, a technique which has become widespread in the field of organization theory (Carley & Prietula, 1994; Prietula et al., 1998). These models consist of three complementary components: the task model, which represents the problem being met by the organization; the model of individual actors, which represents how human actors operate individually; and the model of the organization, which represents how human actors interact to collectively dictate the behavior of the organization. As with any abstraction, there is a fundamental trade-off between the fidelity of the model and the practicality of evaluation. The aspects to abstract are specific to the kind of organization being represented as well as which features are of interest to the evaluator. For example, in an office environment between peers, social features such as fairness and reciprocity

are important determinants of actor behavior, while their effect may be negligible in military settings where soldiers are drilled to simply execute their orders.

Once we have a **computational model** of the organization without SPAs, we consider the effects of SPAs on the organization. We do not explicitly represent SPAs in the organization, but instead model their effects as changes in organizational constraints and parameters. For example, the proposed ability of SPAs automating communication between actors (Chalupsky et al., 2001) may be modeled as a reduction in the communication delay between actors from the baseline value without SPAs. This approach avoids directly implementing the abilities of SPAs (a substantial area of research in its own right) as well as providing an opportunity for published experimental results from that field to be included in the organizational model.

As an example of this general methodology, we consider organizations that are configured to achieve a particular large scale task, where there is an existing, known plan for achieving that task. Examples of such scenarios include a disaster response effort or a military operation. These kinds of domains are extremely challenging to the people operating in them, because the environment is highly dynamic with frequently tight windows of execution for tasks, and people have limited access to incomplete information, often with slow and unreliable communication. For example, in a military operation, soldiers on the ground must often rely on **synchronous** radio transmissions through human communications operators, who relay information to commanders and other soldiers participating in the operation, which can result in information being delayed or distorted. The same characteristics that make these domains so challenging to current organizations also present a number of promising opportunities for SPAs to improve organizational performance by addressing these limitations.

In these domains, actors typically execute hierarchical plans. In our work, we represent an organization of human actors that execute a hierarchical plan similar to hierarchical task networks such as TAEMS (Lesser et al., 2004). Note that a “hierarchical plan” denotes different levels of plan abstraction and does not necessarily imply that the plan will be executed by a hierarchical organization. Indeed, in our experimental evaluations, we consider horizontal and distributed organizational structures executing these plans, in addition to traditional hierarchical organizational structures.

Under these plans, the actors perform a variety of heterogeneous tasks and communicate in order to make decisions. The basic structure is that the organization must make a number of decisions, represented as decision tasks. Each decision requires a specific set of information as input, and this information is acquired by successfully executing sensing tasks. This abstraction describes a wide range of real-life organizations, from military units that detect and eliminate enemy threats, to businesses that identify customer needs and reallocate investments. In this chapter we focus specifically on abstractions of domains such as military operations or disaster response scenarios.

There are a number of assumptions that govern the behavior of actors in this model. Each task is assigned statically to a single actor, who is responsible for performing that task within a fixed window of time, which is known to the actors. While actors may be assigned many tasks with overlapping execution windows, the actors are limited in the number of tasks that they can perform simultaneously, reflecting well-known cognitive limitations of humans. Each task has a fixed duration, and that once begun, the task must

be performed for its entire duration without interruption. Attempts to perform a task may fail, with the failure occurring randomly during the duration of execution, and a failed task cannot be attempted again.

In order to reduce the effect of task failures on plan success, there are redundant, pre-determined contingent tasks (or contingencies) built into the plan. Non-contingent tasks (both sensing and decision tasks) are termed primary tasks, with a specific, known contingency that is to be executed if the primary task fails. Contingencies can also provide redundancy for other contingencies, so that for each primary task, there may be a sequence of contingencies to be invoked in order, in case of successive task failures. Such a sequence of tasks (beginning with the primary task) is termed a task chain, and for each task chain the organization has as a subgoal that one task be successfully completed in that chain, with preference given for tasks earlier in the chain. For each task chain, the responsibility for ensuring that the subgoal is achieved is invested in a manager role, which is statically assigned to one of actors. Managers have the sole power of authorizing the execution of contingencies.

Actors gain information through the successful completion of sensing tasks. This information must then be communicated to the decision makers that use the information. An actor assigned a decision task can only perform that task if it has all the required information. Actors can only transmit information along predetermined, bidirectional communication links, and are bounded in the number of communication links they can maintain. Communication is assumed to be point-to-point, multi-hop with each hop taking a fixed amount of time.

We do not explicitly represent SPAs in the organization, but instead model their effects as changes in organizational constraints and parameters. For example, the presence of SPAs may be modeled by decreasing the communication delay from a baseline for an organization without SPAs. In the rest of this section we formally model an organization and the human actors that comprise it.

Formally, a human organization O is a tuple

$$O = \langle A, T, I, Prov, Req, Window, Dur, Cap, Fail, G, delay_c, M, d \rangle$$

with the following components:

- $A = \{ a_1, a_2, \dots, a_m \}$: the set of *human actors* in the organization.
- T : the set of tasks to be performed by the actors. Tasks are partitioned in two orthogonal ways:
 - $T = T_P \cup T_C$
 - $T_P = \{ t_1, t_2, \dots, t_n \}$: the set of *primary tasks* to be performed.
 - $T_C = \bigcup_{t_i \in T_P} C_i$: the set of *contingencies* to be performed in case of task failures.
 - $T = T_D \cup T_S$
 - T_D : the set of *decision-making tasks* that require information to be performed.
 - T_S : the set of *sensing tasks* that provide information when successfully performed.
- $I = \{ I_1, I_2, \dots, I_l \}$: the set of information produced by sensing tasks and used by decision tasks.

- *Prov*: $T \rightarrow P(I)$: the information provision function mapping (sensing) tasks to the set of information provided by successfully completing the (sensing) task.
- *Req*: $T \rightarrow P(I)$: the information requirement function mapping (decision-making) tasks to the set of information required to perform the decision-making task.
- *Window*: $T \rightarrow [0, \infty) \times [0, \infty)$: the execution window function mapping tasks to the earliest and latest times that the task can be performed.
- *Dur*: $T \rightarrow [0, \infty)$: the duration function indicating how long each task takes to perform.
- *Load*: $T \rightarrow [0, \infty)$: the load function indicating how much **cognitive load** each task requires.
- *Cap*: $A \times T \rightarrow [0, 1]$: the capability function indicating how well actors perform each task.
- *Fail*: $A \times T \rightarrow [0, 1]$: the failure function indicating the probability actors will fail at performing each task when attempted.
- *G*: the undirected graph with actors as vertices and edges indicating possible communication paths between actors.
- *delay_c*: the communication delay.
- *M*: the assignment matrix of primary tasks and contingencies to actors. Each primary task or contingency must be assigned to exactly one actor.
- *d*: the deadline by which time all decisions must be made in order for the organization to succeed.

For each primary task $t_i \in T_P$, there exists 0 or more contingencies, $C_i = \{c_i^1, c_i^2, \dots, c_i^{k_i}\}$. The contingencies are pre-determined backup tasks that are invoked in case of task failure, and are ordered by preference. Hence contingency c_i^1 must be performed if primary task t_i fails, and contingency c_i^2 must be performed if contingency c_i^1 fails, and so on. The set of all contingency tasks is written as T_C .

Several of the model components represent the abilities and limitations of the actors, although there is no explicit model cognitive model of the human actors. The abilities of the actors are represented by the capability function, *Cap*, which incorporates and abstracts the many features of task execution to provide a numerical score for how well each actor performs each task. This function incorporates both the cognitive and physical abilities of the actors used in performing tasks. The *Dur*, *Load*, and *Fail* functions represent some of the limitations of the actors.

The results obtained using this model will be dependent on the fidelity of the model to the actual organization being represented. One point of concern is the parameters (such as *Cap*, *Dur*, etc.) that represent individual actor abilities. There are several ways to determine values for these parameters. Case studies of similar organizations, if available, can be used to supply the values directly. Models of human cognition, such as ACT-R (Anderson, 1993; Byrne, 1999) or SOAR (Laird et al., 1987), can also be used to derive the parameter values for individual actors. Because the organizational model does not require a specific cognitive model of the actors, any available or applicable such model may be used to determine the parameter values. When the specific details of the organization are not known (or when a general class of organizations are under consideration), a range of parameter values should be considered.

Organizational Structure

The undirected graph G in the specification of the organization describes the structural communication links that exist between members of the organization. The vertices of G are the actors in A . An edge between $a_i \in A$ and $a_j \in A$ indicates that a_i can communicate with a_j . There are two cases where an actor must communicate with another actor. The first case is when an actor supplies information to another actor, i.e., a path must exist between two actors a_i and a_j if $M(a_i, t) = 1$ and $M(a_j, t') = 1$ and $Prov(t) \cap Req(t') \neq \emptyset$. The second case occurs when an actor fails at its task, and the fact of this failure must be communicated to the manager of the task chain, who then authorizes execution of the contingency.

Because many contingencies may not actually be performed, some of the edges in G (namely those which only transmit information for those tasks) may not actually be used. Moreover, even those edges that are on paths required for primary tasks may not be used if the actor performing the task fails, since a contingency would then be invoked. Hence we distinguish between two graphs, G and G' . We refer to G as the potential structure graph, since it must contain all edges that could be required for any possible combination of tasks and contingencies. G' , on the other hand, is called the instantiated structure, and only contains those edges that are actually being used given the current set of tasks and contingencies.

Human cognitive limitations restrict the number of contacts a person can effectively maintain. We model this as constraints on the potential structure graph G . In particular, we bound the degrees of the actors in G such that for all $a \in A$, $degree(a) \leq max_{deg}$. Cognitive limitations also impose constraints on the structure of G' , given by max'_{deg} . As tasks are executed and contingencies invoked, the links in G' change. If the degree of an actor in G' exceeds its maximum bound, that actor is said to experience **communication overload**, and is unable to use the communication links that exceed the bounds. Here we assume that the bound is $max'_{deg} = 1$, which models the standard **synchronous** communication prevalent in human organizations today: actors can only communicate with a single other actor at a time, and both parties must be engaged in conversation in order for communication to occur.

For this domain we chose to model the organizational structure as a graph of the possible communication links in the organization. While this modeling choice was motivated by the specific domain of interest, the approach of using a graph to represent possible actions between organization members is quite general and powerful. For more information, please see chapter 10 “Structural aspects of organizations” by Davide Grossi and Frank Dignum.

Execution

The actors attempt to perform primary sensing tasks assigned to them in M . For any primary task failure, contingencies are invoked as described earlier. If an actor succeeds at a sensing task t_s (whether it is a primary task or contingency), the actor gains information $Prov(t_s) \subseteq I$. This information has an associated quality which we denote by $Qual: I \rightarrow [0, 1]$. For all $I_j \in Prov(t_s)$, we let $Qual(I_j) = Cap(a, t_s)$.

The actor then transmits the newly learned information to all neighbors as defined in G . These neighbors store the information in their personal knowledge bases and also in turn propagate the information to their neighbors.

When an actor assigned a decision-making task gains all the information required to make the decision, either from other actors or by performing the sensing tasks itself, it attempts to perform the decision (which can fail). If the actor succeeds, the decision is made with a quality that depends on both the capability of the decision-making actor and the quality of the provided information, and the organization then gets a reward for making the decision. This reward is denoted $Reward: T_P \cap T_D \rightarrow \mathbb{R}$ and is defined in the following way. For a decision-making task t_d successfully performed by actor a , reward is given by:

$$Reward(t_d) = Cap(a, t_d) \left(\sum_{I_j \in Req(t_d)} Qual(I_j) \right)$$

where $Qual(I_j)$ is as defined above. The total reward earned by the organization will be one measure of organizational performance, as described next.

Organizational Performance

The performance of an organization is complex and multi-attributed. There are many possible measures of performance, such as privacy (Maheswaran et al., 2005), resource usage (Lenox et al., 2000b), and volume of communication (Chalupsky et al., 2001). In this work, we focus on the following three specific measures.

- *Success rate*. This is a measure of how well the organization handles individual failure and is calculated as the fraction of the time the organization makes all the decisions represented in T_D before the deadline d .
- *Reward*. This is a measure of how well the organization makes decisions when it succeeds, and is calculated as

$$\sum_{t_d \in T_D \cap T_P} Reward(t_d)$$

- *Speed*. This is a measure of the organization's efficiency and is calculated as the difference between the deadline time d and the time at which the last task finishes execution.

The overall objective function used for performance is a weighted sum of these three measures.

MODELING PERSONAL ASSISTANTS

The SPAs are not directly represented in our model. Instead we model their effects on the constraints of the human actors in the model. We do not model aspects such as social comparison or reciprocity as motivating factors for the SPAs, because these are concepts from agent societies between peers, while the SPAs in our organizations exist solely to assist their human users and are subservient to them. The abilities of SPAs are generally being designed to overcome perceived cognitive limitations of humans that are thought to limit the performance of organizations. In this section, we describe three such limitations

and the models of SPA abilities --- communication management, contingency management, and decision support --- that might overcome these limitations.

Many other SPA abilities (and corresponding cognitive limits) might have been modeled. We focused on this initial set because there was active research in the field on developing these abilities. SPA abilities related to communication management include processing incoming e-mail (Maes, 1994) and asynchronously conveying information previously transmitted by walkie-talkie (Wagner et al., 2004). Transferring responsibility of tasks (Wagner et al., 2004; Schurr et al., 2005) is an SPA ability related to contingency management. Integrating dynamic information and providing users with suggestions (Mitchell et al., 1994; Lenox et al., 2000b) has been studied as a way of SPAs providing decision support to users. This research has focused on developing the SPA abilities and not on the effect of the SPA abilities on organization performance. Furthermore, the question of how the organization can change to better utilize the abilities has not been addressed.

Our choice of selected SPA abilities was also restricted to those most suited to the special purpose organizations that are the focus of this chapter. Future work will expand the scope of SPA abilities, including those where there is not yet active research.

Communication Management

Utilizing SPAs to improve communication is one of the most obvious and heavily researched SPA abilities (Maes, 1994; Malone, 1987). The rate at which decision-makers in an organization can effectively make decisions is limited by two key factors: (i) their ability to get the appropriate information to make required decisions and (ii) their innate processing speed for making those decisions. While wired and wireless networks can convey information at amazingly fast rates compared to human communication speeds, in most cases people are still required to input information and subsequently process received information. The input and output times are subject to human limitations and are fast becoming (if they are not already) a major bottleneck in intra-organization communication, one which could be alleviated by SPAs.

The cognitive load of managing incoming information and directing outgoing information is also a significant constraint that may be relaxed by SPA technology. Specifically, a decision-maker may be able to handle input from a greater number of members of an organization and hence, either get more direct access to required information or make more decisions. For example, the CEO of an organization needs information from many parts of the organization to make key strategic decisions. However, it is cognitively and organizationally infeasible for a person to directly receive input from a large number of people (Carley, 1995). Hence, in the case of the CEO, that information must pass through a small number of department heads, both delaying the information and potentially distorting it. Conversely, suppliers of information are limited in how many others they can provide information to, requiring organizational structures that channel information and introduce delays and distortion. SPA technology could allow information suppliers to more directly communicate with the decision-makers requiring their information.

We modeled three distinct effects that SPA-enhanced communication could have. Firstly, SPAs could directly reduce communication delay by relieving input and output

transmission bottlenecks. This is modeled as a reduction in communication delay. One effect of this change in the organization is that it may reduce the frequency of **communication overload** in **synchronous communication**, because active links are present in G' for less time. The second effect SPAs could have is in providing **asynchronous** communication between actors, modeled as setting $max'_{deg} = max_{deg}$. This has a major effect in eliminating **communication overload**, even if actual communication delay does not change. The third effect that SPAs could have is in reducing the **cognitive load** of managing incoming information from others in the organization and directing outgoing information to others, so that actors can maintain a greater number of communication links. The effect of this directly impacts the possible **organizational structures** and is modeled by increasing the value of max_{deg} .

Task Contingency Management

In addition to reducing communication overload, an **SPA** may manage contingencies in order to prevent its actor from being cognitively overloaded by tasks. Instead of invoking contingencies only when a domain-level failure is suffered, the SPA may monitor the actor's current status and workload and automatically invoke contingencies for any tasks that would overload the actor. This monitoring can be done in a variety of ways, for example by tracking the user's position via GPS or other method (Chalupsky et al., 2001; Wagner, et al. 2004). We assume that actors schedule their pending tasks in order to avoid overload if at all possible, deferring tasks with the most time remaining. When a task can no longer be deferred (due to impending deadlines, for example), the actor invokes the contingency. While this gives maximum flexibility to the actor, it also reduces flexibility for the contingent actor, because it has less time to execute the contingency.

With SPA-managed task contingencies, an actor's SPA automatically invokes contingencies as soon as a potential overload is detected. This gives the contingent actor more flexibility, but may invoke contingencies that did not have to be invoked, which can lead to reduced redundancy in the plan. Thus SPAs also monitor whether their user can subsequently perform tasks for which contingencies were previously invoked; if this occurs, the SPAs transfer responsibility for the task back to the user.

Similar approaches have been developed or proposed for SPAs in other environments (Chalupsky, 2001; Decker & Sycara, 1997a; Maes, 1994). Intuitively, these SPA capabilities should make the organization more robust to individual failure. Managing contingencies or assigning roles has been a key area of research, but this is the first work that attempts to quantify the benefit of such a capability on an organization and understand how the organization may change to leverage this capability.

Decision Support

As introduced above, a key cognitive constraint on humans in an organization is the number of decisions that can be effectively made in a limited amount of time. **SPAs** could relax this constraint in a variety of ways, e.g., taking over routine tasks to provide more time for decisions (Dent, 1992; Garrido, 1996; Schurr 2005), or retrieving, collating, and

presenting information in a way that speeds up the actual decision-making process (Decker, 1997b; Maes, 1994; Lenox et al., 2000b). Without reference to specifically how it is done, in the computational model of the organization, we consider two possible effects of SPA decision support. In the first possible effect, SPAs simply allow a person to effectively make more decisions in the same period of time by decreasing the duration of decision tasks. In the second possible effect, SPAs reduce the **cognitive load** of each decision task, allowing a person to make more decisions simultaneously, without affecting the duration of any single decision task.

EXPERIMENTAL SETUP

In the experiments presented here we were primarily interested in domains such as disaster response scenarios or military operations in which a large number of actors are deployed in a dynamic environment, uncertain environment. Many of the choices for organizational constraints reflect this interest. For example, in the chaos following a natural disaster, many tasks will have fairly high failure rates and small execution windows, as buildings continue to burn and collapse, etc. In addition, communication will be dependent on synchronous radio transmissions relayed between people with fairly long delays to obtain or transmit information. In such an environment, the people executing the tasks will generally not have the time or resources to find other recourses if the communication infrastructure breaks down due to overload.

We evaluated a large number of organizations with a variety of organizational structures through simulation. The basic experimental setup was as follows:

- 80 human actors.
- 176 total primary tasks and contingencies.
 - 16 primary decision tasks, each requiring information provided by 3 distinct sensing tasks.
 - Each decision task has 1 contingency.
 - Each sensing task has 2 contingencies.

We tested three basic types of **organizational structures**: regular hierarchies, rings, and random structures. Regular hierarchies are obviously hierarchical structures, but the rings and random structures were inherently horizontal organizational structures. In a regular hierarchy, the actors form a complete n -ary rooted tree, where n is the span of control. We tested hierarchies with spans of control of two, three, and four, denoted by RH2, RH3, and RH4, respectively. Every actor in a regular hierarchy has one upward link and a fixed number of downward links, so this structure is invariant to changes in max_{deg} . Note, however, that for a regular hierarchy with span of control n to be allowed at all, max_{deg} must be greater than or equal to $n+1$.

Regular hierarchies are used as representatives of traditional hierarchical structures common to human organizations, and as such, there are several additional constraints on the operation of an organization with such a structure. Firstly, information must always flow upward through the hierarchy, so that decision makers are always ancestors of the actors with sensing tasks. This greatly constrains the assignment of tasks based on the structure. Secondly, control flows downward, so that managers are always ancestors of the actors assigned to the tasks being managed.

In a ring structure, the actors form a cycle. If actors can link to more than two other actors (i.e., $max_{deg} > 2$), actors also form links at regular intervals across the ring to minimize the average path length to every actor. In the extreme, this leads to G being fully connected (when max_{deg}). We consider rings with actors having three, four, and five neighbors, denoted by Ring3, Ring4, Ring5, respectively. Ring structures are used as representatives of flat, peer-to-peer organization.

The third type of structure considered is the non-hierarchical random graph. Links were added randomly between actors so that each actor has a fixed number of communication links, if possible. We considered random graphs where each actor has three, four, and five neighbors, denoted by Rand3, Rand4, and Rand5, respectively. Like the ring, this is envisioned as a flat, peer-to-peer organizational structure, and is included primarily as a reference for the other structures.

The organizations tested consisted of two parts, the organizational structure and the task assignment. We generated a large number of organizational structures and used different task assignment algorithms on each of the structures in order to generate the organizations we used for evaluation. Because the task assignment problem is NP-hard in general, we used a heuristic algorithm to assign the tasks for regular hierarchies and ring structures. This algorithm greedily assigns tasks to actors while balancing the assigned load on each actor so that the chance of cognitive overload is reduced. In addition, care is taken to ensure that information from sensing tasks can reach the decision makers, with preference given so that the distance the information must travel (and hence the delay incurred) is minimized, if possible. The additional requirements of regular hierarchies described above impose additional constraints on the task assignment. A completely random allocation was used for random graph structures.

For each organization generated (i.e., organizational structure coupled with task assignment), we simulated the operation of the organization 100 times. The fraction of simulated runs in which the organization successfully completed the plan was used as an approximation of the success rate. The average reward and speed over successful runs were used as approximations to the organization's reward and speed, respectively. If an organization did not succeed at all in the 100 runs, all three metrics received a score of 0.

The next section will describe the experimental results of the baseline performance without SPAs and performance with SPAs modeled as various relaxations of the baseline constraints. The baseline configuration used the following settings:

- Sensing task durations chosen randomly as 10 ± 1 for primary tasks, 15 ± 3 for the first contingency, and 20 ± 5 for the second contingency.
- Decision durations were chosen randomly as 5 ± 1 .
- Sensing task windows were $[0, 50]$ for the primary task and first contingency, and $[25, 100]$ for the second contingency.
- Decision task windows were $[0, 100]$.
- Probability of failure was chosen uniformly at random from $[0.01, 0.40]$ for each task.
- Deadline: $d = 100.0$.
- Communication delay: $delay_c = 2.0$.
- $max_{deg} = 4$. This setting allows the RH2, RH3, Ring3, and Ring4 organizational structures in the baseline configuration.
- Synchronous communication: $max'_{deg} = 1$.

EXPERIMENTAL RESULTS

The Figures in this section show *performance averages* for different organizational structures. These give a rough measure for how well organizations of that structure tend to perform, and are calculated in the following way. The success rate performance average is the average of the success rates for all organizations of that structure and assignment algorithm. The reward and speed performance averages were calculated as the averages of the reward performance and speed performance, respectively, of all organizations of that structure and assignment algorithm with non-zero success rate. Hence the reward and speed performance averages include the performance of only those organizations that succeed at least some of the time. This indicates how much reward organizations receive or how quickly organizations complete the plan when they succeed.

Baseline Results

The baseline raw performance averages for the RH2, RH3, Ring3, Ring4, Rand3, and Rand4 structures are shown in Figure 1. The RH4, Ring5, and Rand5 structures are *not* presented here, as they violate the baseline constraint $max_{deg} = 4$, but they will be presented for comparison below with the results of SPA-managed communication. In Figure 1, the different organizational structures are plotted on the x -axis, while the y -axis shows the raw, normalized performance averages for each of the three components. The reward performance average, shown in red, is greater for ring structures than for regular hierarchies. This is because the additional constraints on task assignments in the regular hierarchies restrict the pool of actors that can be assigned the task, resulting in less capable agents being assigned. Organizations with the RH2 structure also have a greater reward performance average than those with the RH3 structure, because the same constraints become tighter as the span of control increases. However, even the regular hierarchies have greater reward than the random graphs, showing that they are able to assign tasks to agents with above average capability.

The bottom component, which is not visible for Ring3, Ring4, and Rand3, is the success rate performance average. The Figure indicates that the success rate is low for regular hierarchies (with less than a ten percent success rate for RH2 and RH3) and random structures (less than a ten percent success rate for Rand4), and zero for the ring structure. One possible reason for this is suggested by the small speed performance averages, shown as the top components in the Figure. Such small values indicate that the organizations, when they succeed, do so very close to the deadline. Most of the time, the organization exceeds the deadline and hence fails the plan.

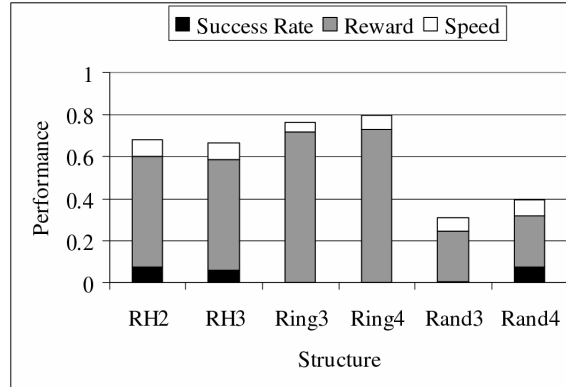


Figure 1: Baseline performance averages.

The organizations tend to exceed the deadline for different reasons, depending on the structure. In ring structured organizations, the average path length between actors tends to be longer than in regular hierarchies. This results in communication that is so slow that information does not reach decision makers until after the deadline, especially when contingencies are invoked, as they tend to be performed by actors that are farther from the decision maker. In regular hierarchies, the tree structure coupled with the requirement that communication flows upward create a bottleneck in communication as information from successfully completed tasks and failure notices for failed tasks overload the communication capacities of actors further up the hierarchy. This prevents information from reaching decision makers quickly, as well as prevents contingencies from being invoked in a timely manner. The RH2 structure has a slightly higher success rate than the RH3 structure because its smaller span of control makes it less likely that actors suffer **communication overload**. As a result, the RH2 structure also has a greater speed performance average than the RH3 structure. The Rand3 structure suffers many of the same problems as the ring structures, but Rand4 performs surprisingly well because random graphs tend to have short paths between all nodes with multiple redundant paths.

Results of SPA Effects on Communication Management

The effect of reducing the communication delay in the organization through SPAs is shown in Figure 1(b)(c)(d). The communication delay is varied on the x -axis, with the baseline value of $delay_c = 2.0$ on the far right. On the far left of the x -axis is the ideal case of instantaneous communication, when $delay_c = 0$. At this point, there is no delay for communication no matter what distance it must travel, and there is no **communication overload** because communications links are active for effectively no time. Thus the performance of the organizations when $delay_c = 0$ are independent of any communication-related issues.

The impact of different communication delays on the success rate performance averages is shown in Figure 2. While all organizational structures improve as the communication delay is decreased, there is a threshold effect with ring structures where performance increases dramatically, far faster than observed with hierarchies. Ring structured organizations typically fail because communication takes too long; contingencies are not authorized until after their execution windows have passed, and

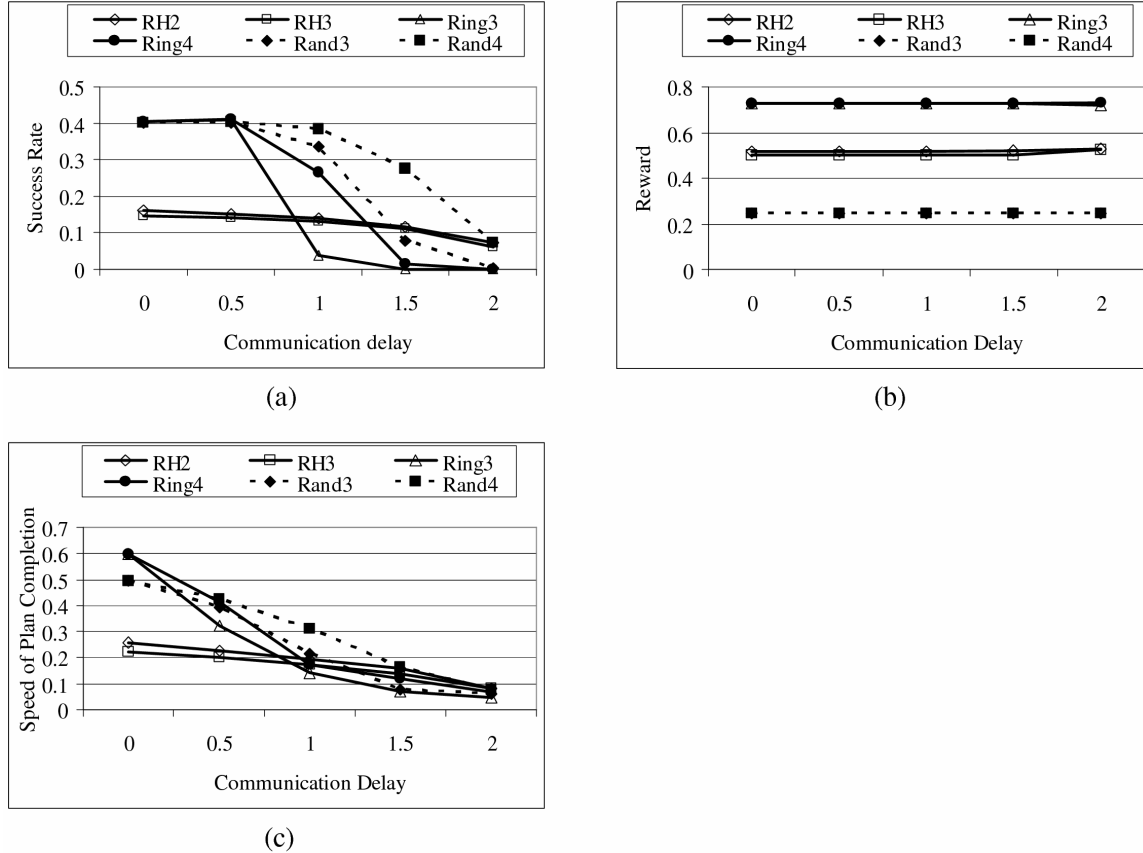


Figure 2: Performance averages with SPAs that reduce communication delay. (a) Success rate. (b) Reward. (c) Speed.

decision makers do not receive required information until after the deadline. The threshold effect occurs because for sufficiently short communication delay, rings are able to communicate in a timely fashion and these failures are avoided. This is further seen in comparing the Ring4 structure to the Ring3 structure. Because Ring4 has a greater number of cross links and hence shorter average paths between actors when compared to Ring3, it also can tolerate higher communication delays, showing a much greater improvement in success rate from a delay of 1.5 to a delay of 1.0 compared to the Ring3 structure, which largely thresholds at a delay between 1.0 and 0.5.

The success rate of the regular hierarchies also improves with reduced communication delay, as can be seen in Figure 2(a). However, it is very interesting to note that regular hierarchies, while starting from much greater success rates than ring structures at the baseline communication delay value, do not improve as much as ring structures do. Even with instantaneous communication shown at the far left of the x -axis, regular hierarchies still succeed only around 15% of the time. This shows that the lower success rates of hierarchical organizations are due to reasons other than communication delay and overload.

The random graphs had higher success rates than the ring structures for moderate values of communication delay. This is because random graphs tend to have shorter paths between actors, which reduces the impact of communication delay.

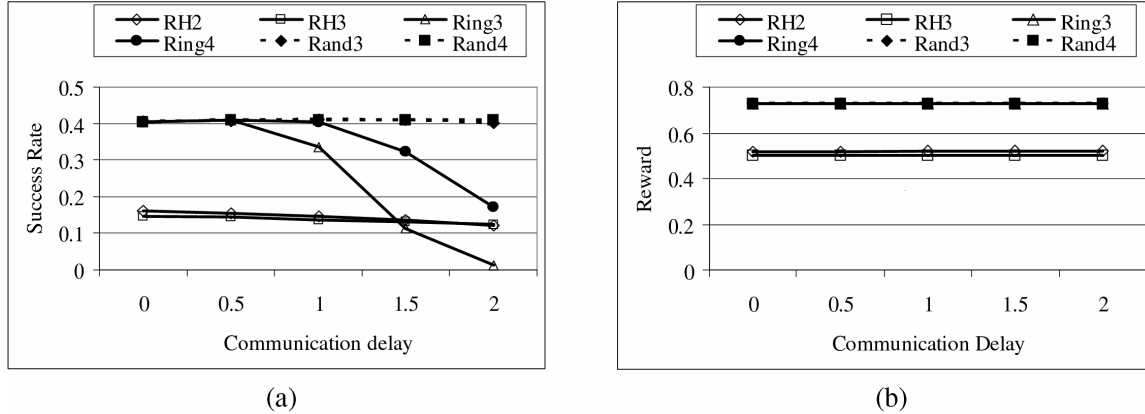


Figure 3: Performance averages with SPAs that provide asynchronous communication and reduced communication delay. (a) Success rate. (b) Speed. Reward performance averages are omitted as they are not significantly different from the synchronous baseline and reduced communication delay cases.

Decreased communication delay has no effect on the reward obtained by organizations, as shown in Figure 2(b). The ring structures clearly obtain more reward when successful than the hierarchical structures, due to the greater flexibility in assigning tasks to ring structures. Both of these structures obtain more reward than the random structures, finding agents of above average capability to execute tasks.

Figure 2(c) shows the effect of communication delay on the speed performance averages. Regular hierarchies improve linearly with decreased communication delay, as the reduced time spent communicating leads to faster plan execution. The relatively low maximum achieved for regular hierarchies when $delay_c = 0$ when compared to that of ring and random structures indicates that even when communication delay and overload are removed, regular hierarchies suffer from other limitations that restrict their ability to quickly perform tasks. The most significant of these (as we will see later) is cognitive overload that arises because the hierarchical constraints on task assignments restrict cause actors to be more heavily loaded than in the ring structure. In contrast, ring structures greatly improve their speed even faster than would be expected by a simple linear speed up, as seen with regular hierarchies and random graphs. The reason for this (as we shall see shortly) is that the shorter communication delay reduces communication overloads, and so task execution is sped up as decision tasks and contingencies that had been waiting for information or authorization at longer communication delays are able to execute earlier.

The effect of asynchronous communication on organizational performance, in conjunction with reduced communication delay, is shown in Figure 3. Figure 3(a) and Figure 3(b) show success rate performance averages and speed performance averages, respectively, with asynchronous communication and varying amounts of communication delay. Reward performance averages with asynchronous communication are unchanged from those seen in the synchronous baseline and reduced communication cases, and are omitted. Because asynchronous communication removes communication overloads, the data shown in Figure 3 indicate the extent to which communication factors other than communication overload affect organizational performance.

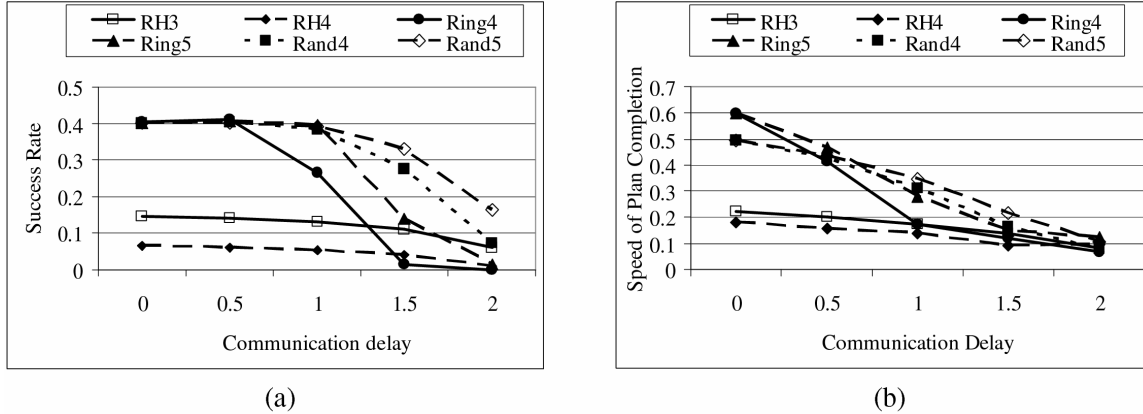


Figure 4: Performance averages for SPAs that increase the number of potential communication links each actor can have. (a) Success rate. (b) Speed.

Figure 3(a) shows that the greatest communication factor limiting regular hierarchies from succeeding are **communication overloads**. At the baseline communication delay, regular hierarchies with asynchronous communication achieve plan success almost twice as often as they do with synchronous communication. Decreasing the communication speed in addition to this also improves the success rate linearly, but the gains that can be obtained by this are not as great as those obtained by using asynchronous communication. This suggests that for SPAs being designed for hierarchical organizations, it may be more beneficial to provide asynchronous communication than to simply increase the rate at which synchronous communication can occur.

Communication overloads are also a significant problem in the ring structures, especially at higher communication delays, as indicated in Figure 3(a) by the greatly improved performances of Ring4 at the baseline communication delay of 2 and Ring3 at delay of 1.5. However, Ring3 continues to perform very poorly with a communication delay of 2, because the long delay still prevents actors from effectively communicating even without communication overloads. Overall, the threshold effect of reduced communication delay on the success rate of ring structures is much less pronounced than it was for synchronous communication. Part of the reason for this is that with synchronous communication, reduced delay affects both communication-related causes of failure, whereas with asynchronous communication, it only affects how quickly information is transmitted in time for decision tasks and contingencies to be executed. From Figures 2(a) and 3(a) it can be seen that the dramatic threshold behavior in synchronous communication occurs primarily when the communication delay has been reduced sufficiently for communication overload no longer to be a major cause of plan failure.

Of all the structures, random graphs are the most affected by **communication overload**. Figure 3(a) shows that the success rate performance averages of random graphs are essentially maximal with asynchronous communication, even for the baseline delay. This strongly suggests that asynchronous communication facilitates flat, decentralized organizational structures.

Figure 3(b) shows that with asynchronous communication, both regular hierarchies and random graphs demonstrate linear speedup with decreased communication delay.

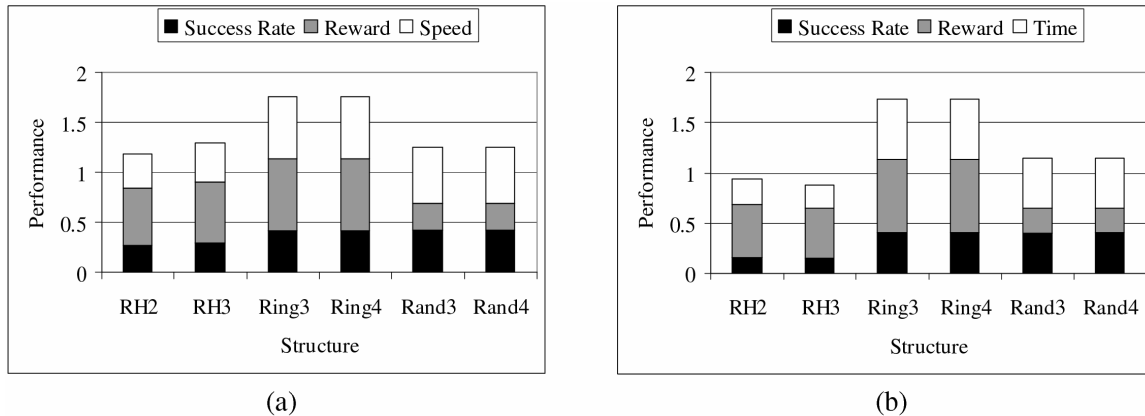


Figure 5: Performance averages (a) with SPA-managed task contingencies; (b) without SPA-managed task contingencies

Ring structures also have a speedup that is much closer to linear than it was with synchronous communication.

Figure 4 shows the effect of using SPAs to manage communications so that the number of communication links each actor can effectively maintain is increased to five. This allows three new organizational structures, a regular hierarchy with span of control 4 (RH4), a ring with 5 neighbors per actor (Ring5), and a random graph with 5 neighbors per actor (Rand5). Figure 4(a) shows the effect on success rate and Figure 4(b) shows the effect on speed of completion. For comparison, each Figure includes the structures with one fewer neighbor per actor (RH3, Ring4, and Rand4). Again, there is no significant change in reward, so that is omitted.

From Figure 4(a) it is clear that adding an additional communication link benefits ring and random structures by allowing faster communication and a greater number of redundant paths, while it harms regular hierarchies because the hierarchical task assignment constraints causes actors to be too heavily loaded to complete all of the requisite tasks before the deadline. Thus even when there is instantaneous communication, RH4 performs very poorly. The effect on speed performance averages in Figure 4(b) are similar to those seen previously, with improvement for ring and random structures and a loss of performance for the regular hierarchy.

Results of SPA Effects on Task Contingency Management

Figure 5(a) shows the performance averages when SPAs allow flexible contingency invocation to prevent cognitive overload. In order to see the potential impact of flexibly and robustly reducing cognitive overload, these results were obtained with a communication delay of 0, so that there were no communication-related causes of failure. For comparison, the performance averages with synchronous communication with delay 0 and no SPA-managed task contingencies is shown in Figure 5(b). From a comparison of Figures 5(a) and 5(b), it is apparent that SPA-managed contingencies greatly increase the success rate performance averages of hierarchical organizations, but have no effect on ring and random structures. This strongly suggests that if SPAs are being developed for a hierarchical organization, task contingency management is one of the most important

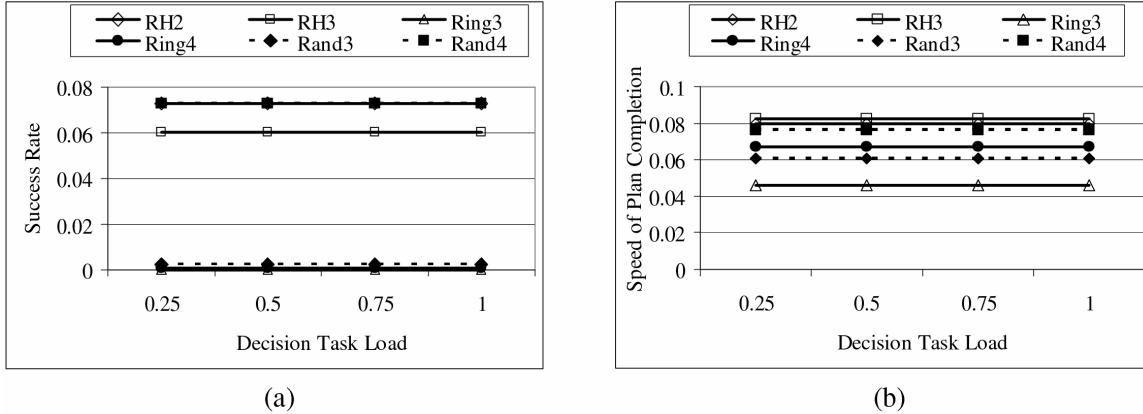


Figure 6: Performance averages with SPAs that reduce decision load. (a) Success rate. (b) Speed.

capabilities to be taken into consideration. However, the performance of the regular hierarchies is still less than that of the ring structures, which suggests that hierarchical organizations with enhanced communication technologies that effectively eliminate communication delay may reap additional benefits by switching to a less centralized organizational structure.

Results of SPA Effects on Decision Support

Figure 6 shows the effect of SPAs that provide decision support by reducing the **cognitive load** of decision tasks. The other organization parameters are set to the baseline configuration. The load of decision tasks is varied on the x -axis. The baseline value of load of 1 is shown on the far right of the x -axis. As can be clearly seen, changing the decision task load has no effect on performance.

Figure 7 shows the effect of SPAs that provide decision support by reducing decision task durations. This has more of an effect on the organizational performance. Figure 7(a) shows that decreasing decision duration linearly increases the success rate of the organization, because faster decisions make it more likely that the organization will meet the deadline. Figure 7(b) shows that decreasing the decision task duration also tends to have a very slight increase on the speed performance average. Because success rates were so low for this setting, there was a larger amount of noise than usual, especially for ring structures.

CONCLUSIONS

Software personal assistants hold great promise to revolutionize human organizations by automating routine tasks and improving coordination between people. Despite various research prototypes and applications, SPA technologies have not yet been able to enter the workplace. We believe that this is due to remaining unsolved technical problems, such as SPAs inferring human intentions, mutual understandability of user and agents, and the unknown impact of introducing SPAs. This chapter is a first step toward addressing this last issue.

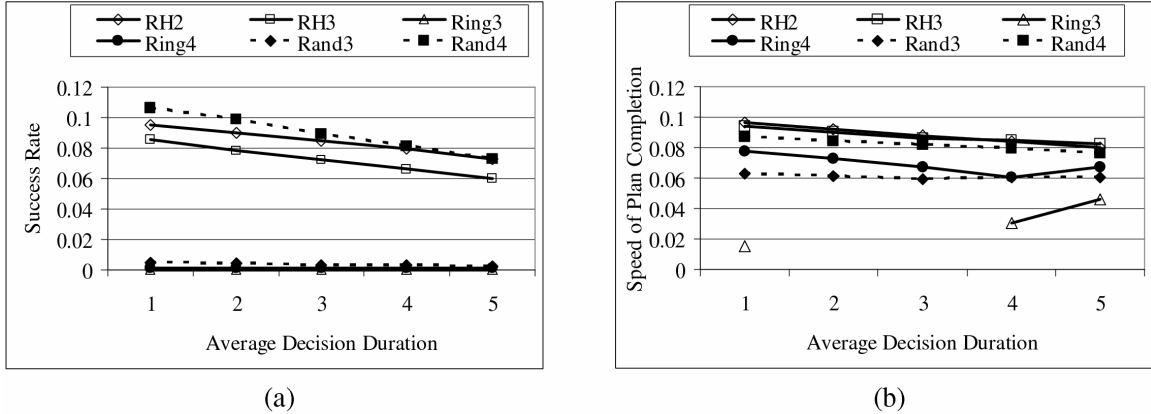


Figure 7: Performance averages with SPAs that reduce decision duration. (a) Success rate. (b) Speed.

In this chapter we developed a conceptual framework and computational model to quantify the impact that various software personal assistant capabilities will have on the performance of human organizations. We found that the type of capabilities that are most beneficial depend on the organizational structure. For hierarchical organizations, SPA capabilities to manage cognitive and communication overload were most important, while for decentralized, flat organizations, SPA technologies that increased the speed of communication were the most helpful. This suggests that SPA designers should consider the appropriate set of SPA technologies depending on the target application, as they would confer the greatest benefits on the organization when deployed. However, the immediate application of SPAs to an organization is only the first phase of their adoption. Our results showing that flat, decentralized organizations perform best overall in the presence of SPAs suggests that over time, even hierarchical organizations may become less centralized in an effort to greater leverage the enhanced communication and coordination offered by SPAs.

ACKNOWLEDGEMENTS

This research was sponsored in part by AFOSR grant FA9550-07-1-0039 and in part by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

Ahuja M. & Carley, K. M. (1998). Network structure in virtual organizations. *Organization Science*, 10(6), 741 – 757.

Ackoff, R. L. (1967). Management misinformation systems. *Management science*, 14(4), B147 – B156.

Anderson, J. (1993). *Rules of the mind*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Byrne, M. D. (1999). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, 12, 439 – 462.

Carley, K. M. & Prietula, M. J. (Ed.). (1994). *Computational Organization Theory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Carley, K. M. (1996). Communicating new ideas: The potential impact of information and telecommunication technology. *Technology in Society*, 18(2), 219 – 230.

Carley, K. M. (1995). Computational and mathematical organization theory: Perspective and directions. *Journal of Computational and Mathematical Organizational Theory*, 1(1), 39 – 56.

Carley, K. M. (2002). Computational organization science: A new frontier. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), 7257 – 7262.

Chalupsky, H., Gil, Y. Knoblock, C., Lerman, K., Oh, J. Pynadath D. V., Russ, T. A., & Tambe, M. (2001). Electric Elves: Applying agent technology to support human organizations. In *Proceedings of the Thirteenth Conference on Innovative Applications of Artificial Intelligence* (pp. 51 – 58). AAAI Press.

Chen, L. & Sycara, K. (1998). WebMate: A personal agent for browsing and searching. In *Proceedings of the 2nd International Conference on Autonomous Agents* (pp. 132 – 139).

Dent, L., Boticario, J., Mitchell, T., Sabowski, D., McDermott, J. (1992). A personal learning apprentice. In William Swartout (Ed.), *Proceedings of the 10th National Conference on Artificial Intelligence – AAAI-92* (pp. 96 – 103). MIT Press.

Decker, K. & Sycara, K. (1997a). Intelligent adaptive information agents. *Journal of Intelligent Information Systems*, 9(3), 239 – 260.

Decker, K., Pannu, A., Sycara, K., & Williamson, M. (1997b). Designing behaviors for information agents. In *Proceedings of the First International Conference on Autonomous Agents* (pp. 404 – 412).

Dignum, M.V., Vázquez-Salceda, J. & Dignum, F.P.M. (2004). OMNI: Introducing social structure, norms and ontologies into agent organizations. In P. Bordini & et al. (Eds.), *Programming Multi-Agent Systems* (pp. 183-200). Heidelberg: Springer.

Garrido, L. & Sycara, K. (1996). Multi-agent meeting scheduling: Preliminary experimental results. In *Proceedings of the Second International Conference on Multi-Agent Systems*.

Garton, L. E. & Wellman, B. (1993). Social impacts of electronic mail in organizations: A review of the research literature. (Tech. Rep. KMDI-HP-93-13). Toronto, Canada: University of Toronto, Knowledge Media Design Institute.

Grossi, D., Dignum, F., & Meyer, J.-J. C. (2007). A formal road from institutional norms to organizational structure. In *Proceedings of the Sixth International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Grossi, D., Dignum, F., Dignum, V., Dastani, M., & Royakkers, L. (2006). Structural aspects of the evaluation of agent organizations. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Gurbaxani, V. & Whang, S. (1991). The impact of information systems on organizations and markets. *Communications of the ACM*, 34(1), 59 – 73.

Horling, B. (2006). *Quantitative organizational modeling and design for multi-agent systems*. Doctoral dissertation, University of Massachusetts at Amherst.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In *Proceedings of CHI '99, ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 159 – 166). ACM Press.

Laird, J. E., Newell, A., Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artificial Intelligence*, 33(1), 1 – 64.

Lenox, T., Hahn, S., Lewis, M., Payne T., & Sycara, K. (2000a). Agent-based aiding for individual and team planning tasks. In *Proceedings of the International Ergonomics Society/Human Factors and Ergonomics Society 2000 Congress*.

Lenox, T., Hahn, S., Lewis, M., Payne, T., & Sycara, K. (2000b). Task characteristics and intelligent aiding. In *Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1123 – 1127).

Lesser, V., Decker, K., Wagner, T., Carver, N., Garvey, A., Horling, B., Neiman, D., Podorozhny, R., Nagendra Prasad, M., Raja, A., Vincent, R., Xuan, P., & Zhang, X. Q. (2004). Evolution of the GPGP/TAEMS domain-independent coordination framework. *Autonomous Agents and Multi-Agent Systems*, 9, 87 – 143.

Levitt, R. E., Cohen, G. P., Kunz, J. C., Nass, C. I., Christiansen, T., & Jin, Y. (1994). The “virtual design” team: Simulating how organization structure and information processing tools affect team performance. In Carley, K. M. & Prietula, M. J. (Eds.), *Computational Organization Theory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Li, C., Giampapa, J., & Sycara, K. (2006). Bilateral contract negotiation decisions with uncertain dynamic outside options. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Special Issue on Game Theoretic Analysis and Stochastic Simulation of Negotiation Agents*, 36(1), 31 – 44.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7), 30 – 40.

Majchrzak, A. & Finley, L. (1995). A practical theory and tool for specifying sociotechnical requirements to achieve organizational effectiveness. In Benders, J. J., De Haan, J., & Bennett, D. (Eds.), *Symbiotic approaches: Work and technology*. London: Taylor and Francis.

Malone, T.W., Yates, J., & Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM*, 30(6), 484 – 497.

Malone, T.W. (1985). *Organizational structure and information technology: Elements of a formal theory* (CISR WP No. 130, Sloan WP No. 1710-85, 90s WP No 85-011). Cambridge, Massachusetts: Massachusetts Institute of Technology, Sloan School of Management, Center for Information Systems Research.

Maheswaran, R. T., Tambe, M., Bowring, E., Pearce, J. P., Varakantham, V. (2004). Taking DCOP to the real world: Efficient complete solutions for distributed event scheduling. In *Proceedings of the Third International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Maheswaran, R. T., Pearce, J. P., Varakantham, P., Bowring, E., Tambe, M. (2005). Valuations of possible states (VPS): A quantitative framework for analysis of privacy loss among collaborative personal assistant agents. In *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

McCrickard, D. S. & Chewar, C. M. (2003). Attuning notification design to user goals and attention costs. *Communications of the ACM*, 46(3), 67 – 72.

Mitchell, T. M., Caruana, R., Freitag, D., McDermott, J., & Zabowski, D. (1994). Experience with a learning personal assistant. *Communications of the ACM*, 37(7), 80 – 91.

Modi, P., Veloso, M., Smith S., Oh, J. (2004). CMRadar: A personal assistant agent for calendar management. In *Proceedings of the Sixth International Workshop on Agent-Oriented Information Systems* (pp. 134 – 148).

Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D., Morley, D., Pfeffer, A., Pollack, M., Tambe, M. (2007). An intelligent personal assistant for task and time management. *AI Magazine*, 2007.

Okamoto, S., Scerri, P., & Sycara, K. (2006). Toward an understanding of the impact of software personal assistants on human organizations. In *AAMAS '06: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 630 – 637). New York: ACM.

Payne, T., Singh, R., & Sycara, K. (2002). Browsing schedules: An agent-based approach to navigating the semantic web. In *Proceedings of the First International Semantic Web Conference* (pp. 469 – 474).

Prietula, M. J., Carley, K. M., & Gasser, L. (Ed.). (1998). *Simulating Organizations*. Menlo Park, California: AAAI Press/The MIT Press.

Schurr, N., Marecki, J., Tambe, M., Scerri, P., Levis, J.P., & Kasinadhuni, N. (2005). The future of disaster response: Humans working with multiagent teams using DEFACTO. In *Proceedings of the AAAI Spring Symposium on Homeland Security*.

Serrano J. & Saugar, S. (2007). Operational semantics of multiagent interactions. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Sycara, K. & Lewis, M. (2004). Integrating agents into human teams. In E. Salas & S.M. Fiore (Eds.), *Team Cognition* (pp. 203 – 233). Erlbaum Publishers.

Sycara, K. & Zeng, D. (1996). Coordination of multiple intelligent software agents. *International Journal of Cooperative Information Systems*, 5(2,3), 181 – 212.

Tambe, M., Bowring, E., Pearce, J. P., Varakantham, P., Scerri, S., Pynadath, D. V. (2006). Electric elves: What went wrong and why. In *Proceedings of the AAAI Spring Symposium on “What Went Wrong and Why”*, 2006.

Varakantham, P., Maheswaran, R., & Tambe, M. (2005). Exploiting belief bounds: Practical POMDPs for personal assistant agents. In *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Vasconcelos, W., Kollingbaum, M. J., & Norman, T. J. (2007). Resolving conflict and inconsistency in norm-regulated virtual organizations. In *Proceedings of the Sixth International Conference on Autonomous Agents and Multiagent Systems*. New York: ACM.

Wagner, T., Phelps, J., Guralnik, V., & VanRiper, R. (2004). COORDINATORS: Coordination managers for first responders. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*. Washington, D.C.: IEEE Computer Society.

Yates, J. & Orlikowski, W. (1992). Genres of organizational communication: A structural approach to studying communication and media," *Academy of Management Review*, 17(2), 299-326.

ADDITIONAL READING

Bowring, E., Tambe, M., & Yokoo, M. (2005). Optimize my schedule but keep it flexible – distributed multi-criteria coordination for personal assistants. *AAAI Spring Symposium on Persistent Assistants: Living and Working with AI*.

Decker, K., Sycara, K., & Williamson, M. (1997). Intelligent adaptive information agents. *Journal of Intelligent Information Systems*, 9, 239 – 260.

He Q., Sycara, K., & Finin, T. (1998). Personal security agent: KQML-based PKI. In *Proceedings of the 2nd International Conference on Autonomous Agents* (pp. 377 – 384). New York; ACM.

Hinds, P. & Kiesler, S. (1995). Communication across boundaries: Work, structure, and use of communication technologies in a large organization. *Organization Science*, 6(4), 373 – 393.

Lenox, T., Hahn, S., Lewis, M., Payne, T. R., Sycara, K. (2000). Task characteristics and intelligent aiding. In *Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1123 – 1127.

Nourbakhsh, I., Lewis, M., Sycara, K., Koes, M., Young, M., & Burlon, S. (2005). Human-robot teaming for search and rescue. *IEEE Pervasive Computing*, 4(1), 72 – 78.

So, Y. P. & Durfee, E. H. (1996). Designing tree-structured organizations for computational agents. *Computational and Mathematical Organization Theory*, 2(3), 219 – 246.

Yates, J., Orlikowski, W. J., & Okamura, K. (1999). Explicit and implicit structuring of genres in electronic communication: Reinforcement and change of social interaction. *Organization Science*, 10(1), 83 – 103.

KEY TERMS & DEFINITIONS

Actor: The human members of the organization. The actors are the entities that are actually executing tasks and are explicitly represented in our model. The actors may be assisted by software personal assistants, which relieve various cognitive and communication constraints.

Cognitive load: The amount of an actor's cognitive facilities that is required to perform a task or set of tasks. People are inherently limited in the amount of cognitive load they can bear, and thus relieving cognitive load is one of the major goals for a successful software personal assistant design.

Communication overload: A person's inability to engage in an excessive number of simultaneous communication acts. Communication overload arises because people have bounded capacities for processing communicated information. This is primarily a problem with synchronous communication.

Decision task: A task that requires information generated by successfully executing sensing tasks. A decision task cannot be executed until all requisite information has been acquired.

Organizational structure: The set of relationships among organizational members constraining their possible actions. In this chapter we consider communication structures that restrict with whom an organization member can communicate. We consider hierarchical, tree-based structures as well as horizontal structures, namely rings and random structures.

Primary task: A task that is not a contingency of any other task. Executing a primary task is the preferred way to achieve an organizational

goal, while contingencies are less preferred but allow the organization to cope with failure.

Sensing task: A task that generates useful information when successfully completed. The generated information can be used in a decision task.

Software personal assistant (SPA): An agent that supports a human user by automating routine individual tasks and facilitating coordination with other members of the organization. The effect of a software personal assistant on its user is to relax various cognitive and communication constraints. To be fully effective, software personal assistants must be aware of the organization in which they are situated.

Task contingency: A task to be executed in the event another task cannot be successfully completed. Each task has at most one, unique and specific contingency that is known in advance. A contingency may in turn have its own contingency to be executed in event of failure. Contingencies represent pre-planned backup or alternative ways to achieve an organizational goal.