An Approach to Quantify Workload in a System of Agents

Richard Stocker[#] Neha Rungta[#] Eric Mercer[+] Franco Raimondi[×] Jon Holbrook[*]
Colleen Cardoza[*] Michael Goodrich[+]
[#]NASA Ames Research Center, Moffett Field CA, USA
[+]Brigham Young University, Provo UT, USA
[×]Middlesex University, London UK
[*]San Jose State University, Moffett Field CA, USA
{richard.s.stocker,neha.s.rungta,jon.holbrook}@nasa.gov, {egm,mike}@cs.byu.edu,
f.raimondi@mdx.ac.uk

ABSTRACT

The role of humans in aviation and other domains continues to shift from manual control to automation monitoring. Studies have found that humans are often poorly suited for monitoring roles, and workload can easily spike in off-nominal situations. Current workload measurement tools, like NASA TLX, use human operators to assess their own workload after using a prototype system. Such measures are used late in the design process and can result in expensive alterations when problems are discovered. Our goal in this work is to provide a quantitative workload measure for use early in the design process. We leverage research in human cognition to define metrics that can measure workload on belief-desire-intentions based multi-agent systems. These measures can alert designers to potential workload issues early in design. We demonstrate the utility of our approach by characterizing quantitative differences in the workload for a single pilot operations model compared to a traditional two pilot model.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human Factors

General Terms

Algorithms, Measurement, Human Factors

Keywords

BDI system; workload computation; reduced flight crew operations

1. INTRODUCTION

In a number of complex and safety-critical scenarios, such as airplanes, health care systems and self-driving cars, the role of humans has shifted from manual control to monitoring of autonomous systems operating together, such as autopilots both in aircraft and in cars, automated collision avoidance systems, etc. However, the increasing complexity of automation makes human monitoring and intervention a difficult task, especially when humans are subject to multiple sources of information, possibly on the same perceptual channels, e.g., two visual inputs, and with different priorities on tasks.

In the area of civil aviation, the focus of the experimental evaluation of this work, Chou *et al.* report that *workload management* is a contributing factor in 23% of 324 studied aircraft accidents [11]. Measuring an operator's mental state presents a range of challenges to researchers despite various proposals including subjective rating scales [21], secondary task performance [13], and psychophysiological measures such as eye movements, heart rate, and respiration [25]). Further, traditional human-in-the-loop (HITL) evaluations such as the NASA Task Load Index (TLX) cannot adequately test performance across the operational envelopes of complex work environments during the design stage of the system development cycle. Therefore predictive models of operator states and performance are needed *at the design stage* to supplement HITL assessments to better ensure the safety of, for instance, future flight deck or air traffic controller workstation designs.

Multi-agent systems (MAS) offer an ideal design abstraction and modeling tool for systems involving both humans and automation, but to the best of our knowledge there is no work that relates them to cognitive models of workload. ACT-R and SOAR and other architectures provide the ability to model actual low-level cognitive processes [2, 28, 29, 30, 31], however, these do not provide the flexibility to describe complex systems and interactions at a higher-level of abstraction. Furthermore, cognitive models are too time consuming to build to be practical [29] or are too simplistic in representing the human [43]. The goal of this work is to model and analyze observable behaviors, activities, decision processes, and goals of agents at a high-level. To this end, we present an approach to use the formal models of beliefs-desires-intentions (BDI) systems with human computer interactions in conjunction with the cognitive concepts of workload. We then compute agent workloads during each time step in the BDI system simulation; the workload measures are referred to as *instantaneous workload.*

The bridge between BDI systems and cognitive workload in this paper is built from cognitive research and takes form in a taxonomy [34] that can be leveraged in a discrete MAS system to measure (i) perceptual workload, (ii) decision workload, and (iii) temporal workload. Perceptual workload rep-

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

resents an agent's ability to process signals on various channels (visual, auditory, or haptic). Decision workload represents the effort and time taken by an agent to arrive at a decision. Temporal workload, refers to the load placed on the agent due to multi-tasking. This paper refines the taxonomy of workload measures in [34] and presents an approach to compute these measures during simulation of a discrete MAS model.

To further refine the taxonomy to MAS systems, we provide a configurable framework to specify weights representing the cognitive effort associated with a specific task, action or decision process in an MAS model. Weights allow the modeler to differentiate the cognitive load between complex tasks such as reading a text while driving and simple tasks such as recognizing a stop sign. The weights are intended to be comparative, i.e., a scale from 0 to n, rather than absolute and would be provided by domain experts who rank cognitive effort on different MAS constructs. These weights are used to compute the instantaneous workload at each time point in the MAS model simulation.

We apply our technique on a discrete time MAS model of a reduced flight crew operation. The instantaneous workload of a single pilot operation (SPO) and that of a traditional two-pilot operation (TPO) during the approach phase of flight is computed. The analysis shows that even in a model where the difference between the SPO and TPO is small, the workload measures can provide interesting insights into the two systems. The problem domain is of great interest as reduced flight crew operations are being considered by airlines and aviation certification authorities as the next stage in civil aviation to reduce operating costs. We believe that our approach to instantaneous workload measure can fill an important need in early design stage evaluation. The contributions of our work are as follows:

- A mapping of workload measures in cognitive models to formal MAS models with discrete time.
- A configurable framework to specify weights representing cognitive effort associated with a specific task, action, or decision process.
- An algorithm to compute instantaneous workload at each time point for a discrete time BDI-based MAS framework—Brahms [12, 50].
- An exploratory study that quantifies instantaneous workload of an SPO as well as a traditional TPO during the approach phase of the flight.

2. COGNITIVE WORKLOAD

2.1 Related Work

The literature on cognitive workload is vast; due to space limitations, here we provide a review only of the most relevant works in the area. The notion of workload started with the work in the early 1900s on time-and-motion studies by Gilbreths [19] and F. W. Taylor [51]. Organizational effects on workload were addressed fairly early in the 20th century by P. Drucker [15, 14] and K. Arrow [3]. T. Sheridan [49, 54] and D. Norman [38] both did seminal work in understanding workload and summarizing state-of-the-art. Several human factors researchers have contributed to understanding workload [35], including methods for assessing workload using secondary tasks [24], NASA TLX [46, 37], or real-time measures [7]. Fundamental work in cognitive psychology [53, 1, 42], including work on task-switching, has produced important insights into workload, as has work on developing cognitive models [9].

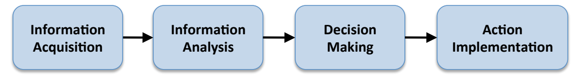


Figure 1: Phases of information processing [41]

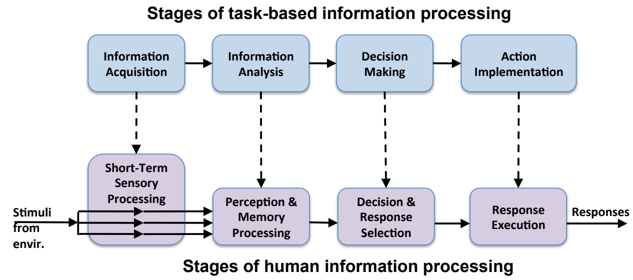


Figure 2: Associating phases of task processing with phases of human information processing.

To the best of our knowledge, the first work to measure instantaneous workload in transition graphs is [34]. Our work refines the taxonomy of workload measures and applies it to MAS frameworks. Other papers that are most relevant to our work include Wickens' foundational and integrative work on multiple resource theory [56] and Parasuraman et al.'s multi-stage model of human information processing [41]. We will make use of each of these papers as we identify key components of cognitive workload that lend themselves to formal modeling.

2.2 Phases of Workload

We want to understand how MAS BDI models can be used to create useful workload predictions for safety critical systems. Toward this goal, it is helpful to identify a high-level framework for how humans and machines process information. We adopt the four-stage model introduced by Parasuraman et al. [41], which is shown in Figure 1.

The main reason for introducing these phases of information processing is that each one involves cognitive resources and, consequently, can impact workload. If we assume that the stages in Figure 1 are for a specific task, then we can identify the stages of human information processing required for that task. More specifically, we can identify specific cognitive resources required to perform a task and, in so-doing, set the foundation for identifying specific elements of a formal model that are associated with specific types of cognitive workload. The mapping from task processing to human information processing is illustrated in Figure 2.

Once an association is made between the phases of performing a task to different cognitive elements, we can elaborate on the interactions among these cognitive elements. This facilitates predicting workload using a formal model by allowing us to identify how specific task elements will induce specific burdens that will, in turn, affect workload. Adapting one of Wickens' models [55] and augmenting it with elements of attention management [32, 48, 42], execu-

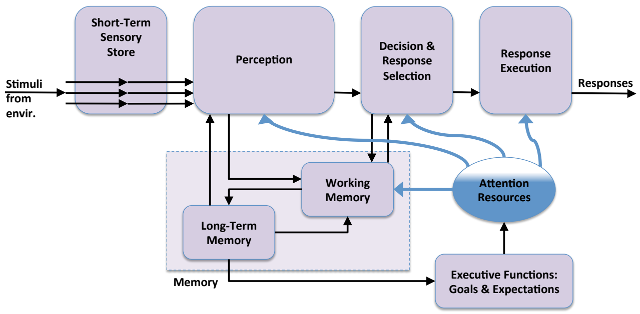


Figure 3: Elaborating on cognitive processes associated with the phases of human information processing (adapted from Wickens [55]).

tive control[1] [10, 33, 4], and working memory [33, 18], yields a slightly-more detailed model for human information processing; see Figure 3.

In 2002, Wickens published a paper on an integrated model of workload along with a process for predicting workload from this model [56]. This model extended the model illustrated in Figure 3 to include a more specific description of the way a human can process multiple signals from the environment. This is illustrated in the figure as the multiple input lines being filtered by short-term sensory processing and being accessed by the perception and memory processes. For our workload models, the key addition is that there are multiple input "channels" through which humans can obtain information. More specifically, there is considerable evidence that humans can sense and perceive signals independently over auditory and visual channels. Note that other aspects of the 2002 model, specifically the different types of coding and the differences between focal and ambient visual attention, are left to future work.

We believe it is possible to translate information-processing models of human cognition into a collection of sub-systems, each with their inputs and outputs. By identifying the inputs and outputs, and then creating models for each subsystem, we can create models that can be used to make predictions about human workload.

2.3 Perceived versus Potential Workload

Before describing our work on using the cognitive model to predict workload, it is useful to discuss the difference between perceived and potential workload. Although high workload can negatively impact performance, a person's perception of workload may not perfectly correlate with their performance [36]. Perception of workload is a strong function of the skill and expertise of the human; experts often perceive the world differently than novices, and may be able to organize signals into "chunks" that do not place heavy burdens on working memory [58, 17].

In this paper, we discuss workload as implicitly defined at the level of expertise instantiated in a model. This means that different models may encode different levels of workload. Consequently, the models discussed in this paper should be understood as *potential* workload rather than workload *perceived* by specific humans. More precisely, in creating a model for a human operator, it is necessary to either explicitly model expertise and training effects or, as we have done in this paper, implicitly account for expertise.

[1]Note that there is an important recent model that avoids some of the issues associated with executive control by replacing this notion with a model of threaded cognition [47].

3. WORKLOAD IN MAS FRAMEWORKS

BDI architectures, originally developed by Michael Bratman [8], are used to represent an agent's mental model of information, motivation and deliberation. Beliefs represent what the agent believes to be true, desires are what the agent aims to achieve, and intentions are how the agent aims to achieve its desires based on its current beliefs. The BDI model is possibly the most well-known and successfully used model for representing rational agents: agents that decide their own actions in a rational and explainable way [45]. This success is perhaps due to its philosophical representation on human practical reasoning processes. This link to human reasoning processes leads to questions on whether metrics of workload, such as the metrics presented in this paper, can be applied to BDI models. In this section we describe the mapping of workload measures to constructs typically found in discrete time MAS models; we also describe how weights on cognitive effort are attached to the MAS model constructs.

The formal model for defining the workload measures is a *Kripke Structure*, $M = (S, S_o, \rightarrow, L)$ (also called *model*), consisting of states (S), initial state ($S_o \in S$), a transition relation ($\rightarrow \subseteq S \times S$), and a labeling function $L : S \mapsto 2^{AP}$ (L returns a member of the power-set over AP) [26]. The set AP contains the atomic propositions that label states in the Kripke structure. Time advances on each transition in the Kripke structure in a fixed quantum to model discrete time semantics. There are no transitions where time does not advance, rather, all the zero-time updates are gathered into a single update that takes place with the advancement of time by the fixed quantum.

Any discrete time MAS model is transformed to such a Kripke structure by finding parts of its reachable state space and then labeling each state appropriately. If the reachable state space advances time at different quantum sizes on any given transition, which is typical of most MAS simulation engines, then the intermediary states and transitions over the smaller fixed quantum must be inserted appropriately. Atomic propositions are organized into partitions to convey information about the original MAS system from which the Kripke structure is derived. In the discussion, $\mathcal{A}$ is the set of agents in the system, and $\alpha \in \mathcal{A}$ is an individual agent. MAS systems typically have the following characteristics:

- **Belief Updates**: constructs whereby an agent updates its belief about the world and the state of other agents. These updates are instantaneous taking zero-time. Labels related to belief updates specific to a given agent, meaning the agent is updating the belief, are in the partition β_α.
- **Activities**: constructs whereby an agent engages in some activity that involves the passage of time such as communication, actuating a mechanical device, or engaging in the real world through movement or other primitive activity. Labels related to activities specific to a given agent, meaning the agent is carrying out the activity, are in the partition Act_α.
- **Tasks**: constructs whereby an agent accomplishes desires and intentions. Tasks consist of belief updates, activities, and sub-tasks to implement complex non-trivial sequencing.

Labels related to tasks that can be performed by the agent are in the partition T_α.
- **Guards**: constructs whereby an agent determines what task to do next, or determines when a current task needs to be suspended, interrupted, or aborted. Labels related to guards specific to a given agent, meaning an agent is evaluation the guard, are in the partition G_α.
The set AP of atomic propositions is defined as $\bigcup_{\alpha \in \mathcal{A}} (\beta_\alpha \cup Act_\alpha \cup T_\alpha \cup G_\alpha)$.

For convenience, the function $\beta_\alpha(s_i) = \beta_\alpha \cap L(s_i)$ returns the belief updates appearing in the label set on state s_i for agent α. Similar functions are defined for activities, tasks, and guards. Also associated with any label in AP is a scaling factor $W : AP \mapsto \mathbb{R}$. The scaling factor is the weight indicating cognitive effort associated with the process represented by a particular construct. A path through the Kripke structure, $\pi = s_o, s_1, s_2, ...$ is a sequence of states such that $\forall i > 0\ (s_{i-1} \rightarrow s_i)$. Each state s_i represents a known point of time in the path.

Instantaneous workload is recorded for each agent at each point in time along a path in the Kripke structure. That workload is defined by a partial function for each agent indexed by an integer i representing the i^{th} time point on the path. As this paper defines three measures for workload, a partial function is associated to each measure for each agent. In the definition of the partial functions, s_i refers to a state in a path of the Kripke structure $\pi = s_o, s_1, s_2, ...$

Three different categories of workload are described based on a small variation of the model in Figure 3.

3.1 Perceptual Workload

Perceptual workload includes two components: multiple salient or relevant signals being broadcast over the same input channel, and the burden placed on working memory to store and integrate necessary signals into a coherent and accurate awareness of the situation. The most common perceptual channels are visual, the things that we see, and auditory, the things that we hear, but haptic and olfactory channels have been used in human factors studies as well. Perceptual workload increases as the required bandwidth over a particular channel increases or when multiple signals are received on an agent's input channel.

In MAS models, agents have to update their beliefs during a simulation run of the model. These belief updates usually happen through perceptual methods implemented in the BDI framework, such as receiving communications or detecting changes in the environment (via sight or sound). Methods by which this is achieved vary from framework to framework but essentially they all map to agent perception. The amount of load in these perceptual channels can be calculated to provide a measure of the agents perceptual workload, i.e., a tally of all perceptions made on each of the agent's perceptual channels (sight, sound, haptic, etc.).

For a given agent, α, β_α are labels associated with zero-time belief updates, and the set Act_α are labels for activities that take time as mentioned previously. Each of these labels is assigned a *type* to indicate the perceptual channel, if any, involved in the update or activity. The function $isP : (\beta \cup Act) \mapsto \{0, 1\}$ is defined using the type of the labels.

$$\text{isP}(l) = \begin{cases} 1 & \text{if } \textit{type}(l) \text{ is related to perception} \\ 0 & \textit{otherwise} \end{cases}$$

Perceptual workload is the number of belief updates or activities at a given time point scaled by their individual cognitive effort. The measure increases when multiple signals are received by an agent.

Definition 3.1 (Perceptual Workload). *The perceptual workload at a state s_i for agent α is*

$$P_\alpha(i) = \sum_{b \in \beta_\alpha(s_i)} (\text{isP}(b) * W(b)) + \sum_{a \in Act_\alpha(s_i)} (\text{isP}(a) * W(a))$$

3.2 Decision Workload

The decision workload encodes the burden placed on working memory and executive control as a function of the difficulty of task selection—choosing what to do next. We assume a fixed level of human expertise, so the measure should be interpreted as a prediction of potential workload. Also note that this definition of decision workload does not account for effective heuristics and approximations that experts may use to make decisions [52].

In BDI based programming languages, such as Brahms [50], AgentSpeak(L) [44], and 3APL [22], agents are able to reason over their current beliefs and formulate a list of actions to perform in order to achieve a goal. Decision workload in a MAS model is the amount of work the agent must perform to formulate this list of actions. The direct measure is the number of reasoning processes the agent goes through in the model and the difficulty of each process indicated by the cognitive effort. In simple terms, the measure counts and weights each guard evaluated in considering a task. A further refinement of the measure would be to account for the amount of working memory required to make these decisions as well as the amount of work required to desired on a final task if multiple tasks are available.

Definition 3.2 (Decision Workload). *The decision workload at a state s_i for agent α is*

$$D_\alpha(i) = \sum_{g \in G_\alpha(s_i)} W(g)$$

3.3 Temporal Workload

The temporal workload encodes the burden placed on memory, attentional resources, executive functions, and interference effects [56, 47] in a multi-tasking context. Temporal workload is represented as a queuing model where service time and arrival rate are used to represent the effort required to manage multiple tasks. Note that this perspective on multi-tasking is at the level of seconds or minutes rather than at the level of milliseconds, with the latter often studied in the cognitive science literature [20].

BDI agents can also produce sets of plans to achieve each of their goals and execute them in a specific order (either randomly or by the priority of the goals they hope to achieve). These sets of plans are often dynamic, meaning the agent can abort the plan or temporarily switch to another and return once the new higher priority plan is finished or no longer has higher priority. This ordering and switching of tasks can be mapped to temporal workload, i.e., the workload assigned to how much the agent needs to multi-task. Temporal workload in MAS models consists of counting the number of tasks the agent is currently executing and weighting each by the associated cognitive effort. For example, if an agent is interrupted by the doorbell and a crying baby while cooking dinner, the temporal workload associated with the agent increases. This measure is best understood as number of ongoing tasks for the agent.

Definition 3.3 (Temporal Workload). *The temporal workload at a state s_i for agent α is*

$$T_\alpha(i) = \sum_{t \in T_\alpha(s_i)} W(t)$$

3.4 A Formal model to quantify Workload

Definition 3.4 (Instantaneous Workload). *The instantaneous workload at state s_i for agent α is*

$$I_\alpha(i) = \zeta * P_\alpha(i) + \eta * D_\alpha(i) + \iota * T_\alpha(i)$$

where ζ, η, and ι are weights to scale the contribution of each category.

In this presentation, $\zeta = \eta = \iota = 1$, but a domain expert may adjust this as well as the other weights for specific systems. Although the rest of this section applies to the Brahms BDI language, the model is general enough to apply equally to any discrete time MAS framework.

3.5 Quantifying workload in Brahms

Brahms is a BDI framework that is designed to model human work processes and robotic activities using rational agents [12, 50]. Brahms' human-centered approach, and its discrete-time event-driven structure make it an ideal choice for demonstrating the workload measures.

Brahms has notions of facts and beliefs, i.e., an agent's interpretation of the world and the actual state of the world using those facts and beliefs. Agents are able to *detect* changes to the environment and update their beliefs, e.g., the agents detect a light flashing on the dash board and update their set of beliefs that the light *is* flashing on the dash board. Beliefs can be exchanged between agents via communications and broadcasts, e.g., the navigation system broadcasts to all in the car that they have arrived at their location.

Belief updates and activities for each agent form the label sets β_α and Act_α respectively. These are further assigned cognitive effort weights and *types* to associate each with a perceptual channel, if any. Perceptual workload is thus computed as in Definition 3.1. The Brahms' simulation labels each state with belief updates or activities that take place at a given time point.

Brahms agents accomplish tasks via constructs known as workframes. A workframe is a guarded event containing a predefined stack of actions and belief updates. A worframe executes only when all its guard conditions satisfied; when more than one workframe can be executed at a time, the priorities of the workframes are used to determine execution order. Within a workframe an agent may perform conceptual activities where only time passes without any update to the state of the agent, communication based activities, location changing activities, and belief updates. Note that Brahms has a notion of top level workframes which can in themselves contain workframes.

Workframes correspond to tasks in the formal model, as such, the label set T_α for a given agent contains all the workframes defined by the agent. Each of these workframes is assigned a cognitive weight. Further, the guards in the workframes for a given agent are put in the set G_α. Each of these guards is also assigned a weight.

Decision workload is computed using the labels in G_α for a given agent. For the Brahms adaptation though, the label is only emitted if the guard evaluates to true. To be clear, when an agent perceives that its state has changed due to an update or activity, then it evaluates it choices of workframes. Any workframe belonging to the agent which has a guard that evaluates to true emits the label for that guard in the state. As such, the decision workload does not account for the cognitive effort in evaluating guards for workframes not related to the current task or workframes with guards that are not enabled in the current state.

Multi-tasking in Brahms is performed by the switching of workframes, either when new higher priority workframes become active or when an agent detects a change in the environment and temporarily switches to another task. Each suspended workframe is kept in a stack until it becomes active again and execution is resumed. Temporal workload in Brahms is measured as in Definition 3.3. The simulation engine emits the task label for each workframe in the stack for a given agent at each point in time in the simulation.

4. IMPLEMENTATION

The humans, automated systems, and objects in a given scenario including their activities and interactions are modeled using Brahms constructs [12, 50]. To quantify workload for the given models, we implement a library to monitor the simulation of a Brahms model and compute the various workload measures within an extensible and customizable Brahms simulation and verification engine [23].

The Brahms compiler generates an XML file that encodes the various data structures of a Brahms model. We extract the labels of the various rules, activities, workframes, etc., of the agents and write them to a configuration file. We automatically assign a default weight of *one* to each of these constructs. The value of one represents a low cognitive effort associated with the construct. This configuration file is then presented to the domain expert for the given model. The domain expert adds the values for the various constructs. Note that perceptual, decision, and temporal workload are weighted equally in this work; this is not based on any cognitive workload theory. We expect the domain expert to weigh the relative importance of each workload category because it may vary from task to task.

We simulate the model along with the assigned weights to generate a set of traces annotated with instantaneous workload measures at each time point. As the constructs with the corresponding labels for a given agent are active, suspended, or being updated during the simulation, the workload computation library tracks the values of the different workload measures. The values are stored, indexed by the time stamp. At the end of the execution, the workload values are written to a CSV file. These results can be used to determine times in the execution that lead to spikes in the workload.

4.1 Driving Scenario

To demonstrate the application of workload metrics to a MAS framework we discuss a simple scenario of an agent driving a car. The activity of driving a car can have varying levels of workload and requires varying degrees of cognitive effort. The model used in our example contains a *car*, *traffic lights*, a *speed limit sign*, an *intersection*, a *navigation sys-*

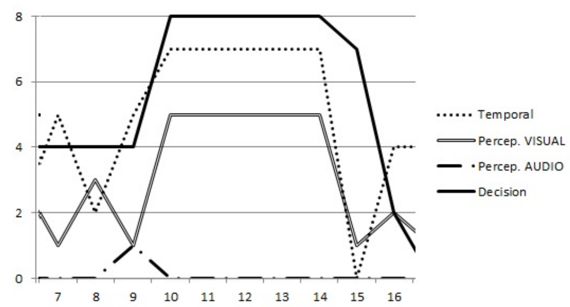


Figure 4: Different workload measures of the driver in a car approaching traffic lights.

tem, a *passenger*, a *cell phone* and the *driver*. During the simulation the driver receives visual signals in the form of changes to the traffic lights, appearances of stops signs, and coming up to an intersection. The driver also receives auditory signals by listening to the radio, to the passenger in the car, and instructions from the navigation system. The driver makes decisions regarding whether to read a text message on the cell phone, check mirrors, check speedometer, stop at the red light, etc. The model allows the driver to multi-task while driving, e.g., listen to navigational instructions as well as read the text message on the cell phone. For this model we have obtained the weights by means of an informal survey among 6 drivers, asking them to assign weights to the cognitive effort associated with various belief updates, activities, events, and tasks in the model. We took the average of the weights and computed instantaneous workload during a simulation. The values of the weights for cognitive efforts are on a 6-point scale from 0 that represented no workload to 5 that represented very high-workload.

Fig. 4 presents the different workload measures of the driver for a small portion of the driving scenario; where the car approaches traffic lights. In Fig. 4, the x-axis represents time points in the simulation while the y-axis represents the workload values. We present measures for the (a) temporal workload, (b) perceptual workload partitioned into the visual channel and the audio channel, and (c) decision workload. The sum of temporal, perceptual, and decision workload values represents the instantaneous workload.

In Fig. 4, between the times seven and nine the agent is driving at a constant speed alternating between watching the road, checking mirrors, and reading the speedometer. At time point nine the agent receives a text message, the sound notification of the text message causes a rise in AUDIO perception workload of the driver. Once the message is received, the driver is faced with a decision on whether to read this message or not; this increases decision workload to 7 at time point ten. In the scenario, the driver chooses to read the message. While reading the message the agent has a fairly high decision workload choosing between whether to stop, continue to watch the road, read speedometer, or check mirrors. The temporal workload throughout this activity is also high since the driver is performing multiple tasks, driving the car as well as reading the text message. The driver finishes reading the message at time 15, where all the workload values temporarily drop. At time point 16 the driver then notices a change in the traffic lights which causes an increase in VISUAL perception workload.

5. REDUCED CREW OPERATION

In the 1950s, a 5-person crew was required to fly large transport category aircraft. Due to advancements in the design and performance of aircraft systems, this number decreased to two pilots in the 1980s. Aircraft technology developers have continued to improve automation, further reducing crew workload associated with manual flying tasks, particularly for the en route / cruise portion of the flight. However, automation has increased demands on the pilots to monitor systems for possible failures, a role for which research has shown that humans are poorly suited [40, 57]. Despite the reduction in crew size and the evolving roles of the pilots, commercial aviation operations remain extremely safe, far exceeding safety levels seen in single-pilot general aviation or air taxi operations [39]. It is unclear, however, the extent to which these higher safety levels are attributable to the presence of an additional pilot, because these types of flight operations differ in many ways, e.g., levels of automation, ground support from ATC, and aircraft complexity.

As flight deck automation continues to evolve, airline operators and manufacturers have begun to explore concepts to further reduce crew size in large transport aircraft to a single pilot. If alternative methods could be found for supporting the remaining high-workload portions of the flight such as takeoff, approach and landing, an SPO could potentially reduce operating costs for the airline operators. Any transition from current the current TPO to SPO, however, must maintain at least current levels of safety while maximizing cost savings. Research into SPO concepts has focused on increasing flight deck automation, adding ground-based assets to support the lone pilot, or, more commonly, some combination of these approaches. See [6] for a review of SPO concepts of operation. In most SPO concepts, a ground operator would support a lone pilot specifically during high-workload phases of a flight, thus allowing a single ground operator to potentially support multiple flights, thereby reducing overall crewing requirements. However, operating as a distributed team can result in communication breakdowns that require additional coordination [5].

In this paper, we explore a method for measuring the potential impact of SPO on human operator workload. We instantiated SPO in our model based on a recent SPO concept that included ground operator support for a lone pilot [27]. We chose to model the "final approach" phase of flight, due to its associated high workload. Final approach describes the last leg of an aircraft's flight path, immediately before landing. The final approach segment typically begins about 5 nautical miles from the arrival end of the runway and at about 3,000 feet above the ground. On final approach, the aircraft is flying along a path aligned with the runway. At the same time, the aircraft is also descending for landing. The final approach segment is a high workload time for pilots because they need to follow a large series of predetermined maneuvers within a relatively short period of time. The aim of these maneuvers is to slow and to configure the aircraft correctly so it can descend on a specific flight trajectory. Final approach concludes with the aircraft landing within the touchdown zone of the runway and at the correct touchdown speed.

5.1 Models

In this section we describe details about the models created to evaluate workload in a reduced crew operations. We create a Brahms model for a TPO approach and quantify its workload in the same setting as SPO to provide a means for comparison. We create Brahms model of an SPO with ground support during approach.

The models were created based on the input from an experienced pilot. The approach is broken up into six stages; approx. 2000 feet, 1500 feet, 1000 feet, 500 feet, 200 feet and 40 feet above the ground. For each of these stages, the pilot provided information about key tasks and actions performed by the pilot, the type of communication taking place in the cockpit, and the weights that determine the cognitive effort on the various constructs.

5.1.1 Two-Pilot Model

The two-pilot scenario is of a nominal approach from five miles out and 2000 feet up. The pilot is the person *flying* the plane, interacts with controls of the plane, reads values off the instruments and communicates with the first-officer who is monitoring. There are no off-nominal conditions possible in the model. There are several activities performed by the pilot that are over the visual and audio perceptual channels, such as communications from the pilot monitoring and reading values off instruments. The pilot is often choosing between a variety of tasks with the same priority such as checking instruments and communicating with the first officer. The pilot can also be required to multi-task on several occasions due to incoming alerts or readings which require the pilot to temporarily switch tasks.

5.1.2 Single Pilot Model

The SPO model is a duplication of the two-pilot model with some minor adjustments. Note that this is by design, our goal was to minimize the number of differences between the SPO model and the two-pilot model. The pilot essentially performs the same tasks in the SPO as in the TPO, but instead of communicating with the first-officer on board, the pilot is communicating with an operator on the ground. We assume that the operator on the ground is able to access the same instrument data that would have been available to the first officer. The ground operator, however, may be lacking certain situational awareness. e.g., the operator on the ground is unable to feel the movement of the plane and may not realize fluctuations in readings due to wind shear. The ground operator may request additional information from the pilot which a first officer in a two-pilot scenario would not.

5.1.3 Workload Measures

Fig. 5(a) and (b)[2] show the workload measures for the six phases in the approach part of the flight in the SPO and TPO models respectively. The workload measures have been averaged across 12 runs. Each phase is demarcated by dashed vertical lines. The x-axis represents the time points in the simulation while the y-axis represents the workload values. The graphs show that the workload increases as the pilot moves through each phase. The exception to this is the decision workload. This is due to the fact that in a nominal approach phase of flight, the pilot is not faced with different choices. It can also be noted that the perceptual workload falls in the final stage in both graphs because the pilot is communicating less and focused on landing the plane.

During the first three phases the workload of the pilot gradually increases in both Fig. 5(a) and (b). In phases four and five, there is a large increase in the perceptual and temporal workload because the pilot has to monitor different instruments very closely, maintain communication with the operator on the ground, communicate with the air traffic controller, and perform activities to decrease the speed and altitude of the plane. In the sixth and final phase, there are no communications between the pilot and operator on the ground which causes the perceptual workload to decrease. The temporal workload rises due to the tasks performed by the pilot to keep the plane on course for landing.

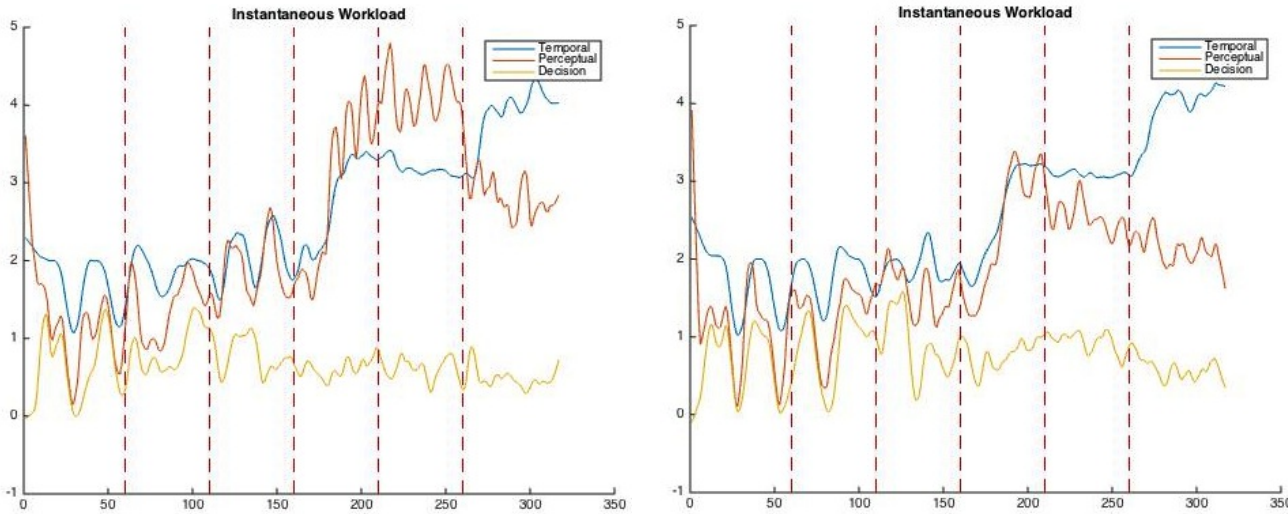
We compare the workload measures of the TPO versus the SPO in Table 1. Table 1(a) presents the average workload values, Table 1(b) presents the average of the maximum observed workload values, Table 1(c) presents the average number of times the workload values reaches the maximum, and Table 1(d) presents the standard deviation from the average values observed in the TPO trials. We present the perceptual workload (Percep.), further divided into the perpetual audio (P-Aud.) and perpetual visual (P-Vis); the decision workload (Decis.); and the temporal workload (Temp.) in each of the tables. In Table 1(a), generally, the average workload values for the SPO are higher compared to that of the TPO, however, the perception across the visual channel is higher in the TPO. This was an unexpected result, an analysis of the data showed this was due to the fact that the pilot had no visual interaction with the pilot not flying in the SPO compared to the TPO. The perception over the audio channel is higher due to the extra communication with the pilot on the ground. This causes the overall perceptual workload to go up in the SPO in Fig. 5(a) compared to the perceptual workload of the TPO in Fig. 5(b). The decision and temporal workload averages are marginally higher. Similar trends are observed in Table 1(b), (c), and (d) and can also be seen in Fig. 5(a) and (b).

The advantage of computing the instantaneous workload, is that it allows us to consider the spikes in the workload and not just average values. The average maximum temporal workload is 5 for both the SPO and TPO as shown in Table 1(b), however, number of times the maximum is reached is 20.6 in the SPO whereas in the TPO it is 16.6 as shown in Table 1(c). This shows that that there are more instances of the temporal workload being high for the SPO compared to the TPO. The same pattern is observed in the decision workload as seen in Table 1(b) and (c). The higher spikes in the workload in the SPO model can be seen in Fig. 5(a).

5.2 Discussion

Modeling outcomes suggest higher perceptual workload for the pilot flying the SPO versus TPO. As described above, perceptual workload includes two components: the processing of salient or relevant signals from the environment over the same sensory input channels, as well as the integration of these signals into existing memory to form a coherent and accurate awareness of the situation. On a two-pilot flight deck, the pilot flying and the first officer (pilot monitoring) perform well-learned procedures to ensure that both pilots maintaine shared awareness of the situation, including ver-

[2]The graphs should be viewed in color.



(a) Single Pilot Operation—SPO

(b) Two Pilot Operation—TPO

Figure 5: Average workload measures during the six phases of an approach across 12 simulation runs.

Table 1: Workload values for the SPO and TPO across 12 runs (a) average of the workload values, (b) average of the maximum observed workload values, (c) average number of times the workload values reaches the maximum, and (d) standard deviation from the observed TPO average values.

Type	Percep.	P-Aud.	P-Vis	Decis.	Temp.
SPO	2.361	0.112	0.65	8.389	2.612
TPO	1.826	0.037	0.772	8.229	2.534

(a)

Type	Percep.	P-Aud.	P-Vis	Decis.	Temp.
SPO	18.538	3.384	3	35.384	5
TPO	13.875	2	3	33.25	5

(b)

Type	Percep.	P-Aud.	P-Vis	Decis.	Temp.
SPO	5.166	12.75	70.25	1.833	20.666
TPO	0.333	2.916	83	0.5	16.666

(c)

Type	Percep.	P-Aud.	P-Vis	Decis.	Temp.
SPO	3.277	0.509	1.241	9.751	1.11
TPO	2.19	0.239	1.309	9.214	1.075

(d)

bal call-outs and physical point-outs of critical information. In addition, pilots observe each other's behavior, and can interpret it within the context of the flight (e.g., the pilot monitoring may observe that the pilot flying is making lots of minor yoke adjustments in response to wind gusts that both pilots are perceiving vestibularly). In the SPO concept we have modeled, however, the ground-based pilot monitoring is unable to use all of these sensory inputs, because the pilots cannot see each other, and the ground operator is unable to perceive minor movements of the aircraft. Therefore, the distributed flight crew in the SPO concept must work harder to maintain shared awareness of the situation through increased verbal communication (i.e., to explicitly describe things that the other crew member cannot see) and through increased monitoring of interpretation of flight instruments (i.e., to infer actions that are being taken by the other crew member and the reasons for those actions).

A necessary condition for the success of SPO is ensuring that single pilot operations are at least as safe as two-pilot operations. Our initial predictions of workload under SPO and TPO suggest that the workload associated with maintaining shared situation awareness would be higher in the evaluated SPO concept. High workload can lead to a decrease in the number of information sources or tasks attended to, incomplete integration of perceived information, divergent mental models of the situation by pilot flying and pilot monitoring, and decreased ability to recognize or cope with unanticipated situations. In a study of major air carrier accidents, 88% of those involving human error were attributable to problems with situation awareness [16]. Therefore, the outcomes of our study suggest that SPO concepts should include mitigation strategies (e.g., new automation, procedures, etc.) to ensure that distributed pilot teams are able to maintain shared situation awareness with the same ease as co-located pilot teams.

6. CONCLUSION AND FUTURE WORK

In this paper we present three measures to quantify workload in multi-agent systems. These measures are leveraged from cognitive psychology and extended for applications in BDI systems and multi-agent systems. We formally define these workload measures and explain how they can be applied to multi-agent systems in general. We present a mapping of these measures to a multi-agent framework called Brahms and demonstrate its utility with a case study comparing the workload of a single-pilot operation against a standard two-pilot operation. As part of future work we plan to develop more complex high workload scenarios in the aviation domain, along with studies with aviation pilots to further validate the accuracy of our workload measures.

REFERENCES

[1] E. M. Altmann and J. G. Trafton. Task interruption: Resumption lag and the role of cues. In *Proceedings of the 26th annual conference of the Cognitive Science Society*, pages 43–48, 2004.

[2] J. R. Anderson, M. Matessa, and C. Lebiere. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462, 1997.

[3] K. J. Arrow. *The Limits of Organization*. W. W. Norton and Company, 1974.

[4] A. Baddeley and S. Della Sala. Working memory and executive control. *Philosophical Transactions: Biological Sciences*, 351(1346):1397–1404, http://dx.doi.org/10.1098/rstb.1996.0123, 1996.

[5] C. Bearman, S. B. Paletz, J. Orasanu, and M. J. Thomas. The breakdown of coordinated decision making in distributed systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(2):173–188, 2010.

[6] K. D. Bilimoria, W. W. Johnson, and P. C. Schutte. Conceptual framework for single pilot operations. HCI-Aero, 2014.

[7] E. R. Boer, T. Futami, T. Nakamura, and O. Nakayama. Development of a steering entropy method for evaluating driver workload. In *SAE 2001 World Congress*, Detroit, Michigan, USA, March 5-8 2001. SAE paper #1999-01-0892.

[8] M. Bratman. Intention, plans, and practical reason. 1987.

[9] M. D. Byrne and A. Kirlik. Using computational cognitive modeling to diagnose possible sources of aviation error. *The international journal of aviation psychology*, 15(2):135–155, 2005.

[10] N. J. Cepeda, A. F. Kramer, and J. C. M. G. de Sather. Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, 37(5):715–730, 2001.

[11] C. Chou, D. Madhavan, and K. Funk. Studies of cockpit task management errors. *The International Journal of Aviation Psychology*, 6(4):307–320, 1996.

[12] W. J. Clancey, P. Sachs, M. Sierhuis, and R. van Hoof. Brahms: Simulating practice for work systems design. *Int. J. Hum.-Comput. Stud.*, 49(6):831–865, Dec. 1998.

[13] D. Damos. Dual-task methodology: Some common problems. *Multiple-task performance*, pages 101–119, 1991.

[14] P. F. Drucker. *The Practice of Management*. Harper Collins, 1954.

[15] P. F. Drucker. The management horizon. *The Journal of Business*, 28(3):155–164, 1955.

[16] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[17] M. R. Endsley. The role of situation awareness in naturalistic decision making. In C. E. Zsambok and G. Klein, editors, *Naturalistic Decision Making*, chapter 26, pages 269–283. Lawrence Erlbaum Associates, Hillsdale, N.J., 1997.

[18] K. A. Ericsson and P. F. Delaney. Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake and P. Shah, editors, *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pages 257–297. Cambridge University Press.

[19] F. B. Gilbreth and L. M. Gilbreth. *Applied motion study: A collection of papers on the efficient method to industrial preparedness*. Macmillan, 1919.

[20] M. A. Goodrich, M. R. Quigley, and K. A. Cosenzo. Task switching and multi-robot teams. In *Proceedings of the 2005 NRL Workshop on Multi-Robot Systems*, pages 185–195, 2005.

[21] S. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[22] K. V. Hindriks, F. S. De Boer, W. Van der Hoek, and J.-J. C. Meyer. Agent programming in 3APL. *Autonomous Agents and Multi-Agent Systems*, 2(4):357–401, 1999.

[23] J. Hunter, F. Raimondi, N. Rungta, and R. Stocker. A synergistic and extensible framework for multi-agent system verification. In M. L. Gini, O. Shehory, T. Ito, and C. M. Jonker, editors, *AAMAS*, pages 869–876. IFAAMAS, 2013.

[24] D. B. Kaber and J. M. Riley. Adaptive automation of a dynamic control task based on secondary task workload measurement. *International journal of cognitive ergonomics*, 3(3):169–187, 1999.

[25] A. F. Kramer. Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, pages 279–328, 1991.

[26] S. A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16(1963):83–94, 1963.

[27] J. Lachter, S. L. Brandt, V. Battiste, S. V. Ligda, M. Matessa, and W. W. Johnson. Toward single pilots operations: Developing a ground station. HCI-Aero, 2014.

[28] J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: An architecture for general intelligence. *Artificial intelligence*, 33(1):1–64, 1987.

[29] C. Lebiere, F. Jentsch, and S. Ososky. Cognitive models of decision making processes for human-robot interaction. In *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pages 285–294. Springer, 2013.

[30] Y. Liu. QN-ACES: Integrating queueing network and ACT-R, CAPS, EPIC, and Soar architectures for multitask cognitive modeling. *International Journal of Human-Computer Interaction*, 25(6):554–581, 2009.

[31] M. Lundin. Simulating the effects of mental workload on performance in tankcrew.

[32] N. Meiran, B. Hommel, U. Bibi, and I. Lev. Consciousness and control in task switching. *Consciousness and Cognition*, 11:10–33, 2002.

[33] A. Miyake and P. Shah, editors. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press, 1999.

[34] J. J. Moore, R. Ivie, T. J. Gledhill, E. Mercer, and M. A. Goodrich. Modeling human workload in unmanned aerial systems. In *AAAI Spring Symposium Series: Formal Verification and Modeling in Human-Machine Systems*, pages 44–49, 2014.

[35] N. Moray, M. J. Dessouky, B. A. Kijowski, , and R. Adapathya. Strategic behavior, workload, and performance in task scheduling. *Human Factors*, 33(6):607–629, 1991.

[36] L. A. Muse, S. G. Harris, and H. S. Feild. Has the inverted-u theory of stress and job performance had a fair test? *Human Performance*, 16(4):349–364, 2003.

[37] NASA Task Load Index (TLX) users manual.

[38] D. A. Norman and D. G. Bobrow. On data-limited and resource-limited processes. *Cognitive psychology*, 7(1):44–64, 1975.

[39] M. Norman. Economic opportunities and technological challenges for reduced crew operations. technical paper prepared for: Crew vehicle interface element, 2007.

[40] R. Parasuraman, R. Molloy, and I. L. Singh. Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3(1):1–23, 1993.

[41] R. Parasuraman, T. B. Sheridan, and C. D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, 30(3):286–297, May 2000.

[42] H. Pashler. *The Psychology of Attention*. MIT Press, Cambridge, Massachusetts, USA, 1998.

[43] A. R. Pritchett and K. M. Feigh. Simulating first-principles models of situated human performance. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on*, pages 144–151. IEEE, 2011.

[44] A. Rao. AgentSpeak (L): BDI agents speak out in a logical computable language. In *Agents Breaking Away*, pages 42–55. Springer, 1996.

[45] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. *KR*, 91:473–484, 1991.

[46] H. E. Resource. NASA-TLX (NASA Task Load Index).

[47] D. D. Salvucci and N. A. Taatgen. Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review*, 115(1):101–130, 2008.

[48] S. Schuch and I. Koch. The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, 29:92–105, 2003.

[49] T. B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, 1992.

[50] M. Sierhuis. *Modeling and Simulating Work Practice. Brahms: a multiagent modeling and simulation language for work system analysis and design*. PhD thesis, Social Science and Informatics (SWI), University of Amsterdam, SIKS Dissertation Series No. 2001-10, Amsterdam, The Netherlands, 2001.

[51] F. W. Taylor. *The principles of scientific management*. Harper, 1914.

[52] G. G. P. M. Todd. *Simple Heuristice That Make Us Smart*. Oxford University Press, New York, 1999.

[53] J. G. Trafton, E. M. Altmann, D. P. Brock, and F. E. Mintz. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58:583–603, 2003.

[54] M. K. Tulga and T. B. Sheridan. Dynamic decisions and work load in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(5):217–232, 1980.

[55] C. D. Wickens. *Engineering psychology and human performance* . HarperCollins Publishers, 1992.

[56] C. D. Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

[57] E. L. Wiener. Application of vigilance research: rare, medium, or well done? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6):725–736, 1987.

[58] C. E. Zsambok and G. Klein, editors. *Naturalistic Decision Making*, Hillsdale, N.J., 1997. Lawrence Erlbaum Associates.