

Natural human role in supervising complex control systems

Henry Hexmoor*, Brian McLaughlan and Gaurav Tuli

Computer Science Department, Southern Illinois University, Carbondale, IL, USA

(Received November 2007; final version received 4 June 2008)

Large collections of communities such as those found in complex control systems necessitate sophisticated techniques for high-level human supervision due to their requirements of influence spanning over individuals, communities, and global system behaviours. We have developed psycho-socio-cultural models for mediation of system-level behaviours and interactions. The ensuing lowered human cognitive load will enable supervisors to effectively guide large systems with competing objectives. This model paves the way for developing the 'Man On The Loop' (MOTL) paradigm, a phrase we are proposing for a novel human supervision role that contrasts with typical micromanagement. We have highlighted and validated MOTL parameters through the implementation of a domain-neutral simulation and compared our results with those found in natural systems.

Keywords: multi-agent system; supervisory control; social-psycho-cultural model; human agent interaction

1. Introduction

A key issue that arises in controlling complex systems is judicious determination of the extent to which human participation is useful. Multifaceted, mission critical systems that are in development for defence and space explorations need to firmly remain under human jurisdiction while the loci of control are increasingly and technologically widely distributed. The level of human intervention may vary from minimal participation to total domination within a control loop. The level of human intervention can be broadly divided into two categories. If human interference directly affects the actions of a specific acting entity (i.e. an agent) in the control loop, it is traditionally termed '*Man In The Loop*' (MITL). Whereas, if the human intervention indirectly affects actions at the unit or agent community level, we coin the phrase '*Man On The Loop*' (MOTL).

Man on the loop is characterised as a paradigm shift where a human supervisor employs psychological and social influences to control the system behaviour. MOTL alleviates excessive dependence on the human supervisor. A system-level network policy management is an example for MOTL, where changing the policies in the policy management system by the system administrator will affect the configuration of the entire network without tediously altering each system parameter (Strassner 2003). Moreover, the human supervisor must attempt to coerce the community to be a

*Corresponding author. Email: hexmoor@cs.siu.edu

responsible society, discharging minimum sets of necessary rights and duties (Jennings and Campos 1997) via:

- Acquaintance: the notion that the society contains other members.
- Influence: the notion that members can affect one another.
- Rights and Duties: the notions that a member can expect certain things from others in the society and that the society could expect reciprocation.

Since decision making in most complex systems is cognitive and deterministic, and validation of such systems is difficult and tedious, MITL has been the predominant paradigm for such systems. This article posits that there is a need to relinquish the intimate, micro-managerial control style of the MITL paradigm to the higher level MOTL paradigm that prescribes a naturalistic, psycho-social macro-management role to human supervisors in charge of the systems. In the MOTL paradigm, a human supervisor will influence individuals and communities of system control units by managing the manner in which control units relate to one another while preserving mission effectiveness and performance metrics. However, it must be stressed that successful task performance invariably depends on the capabilities of the users (McFarlane and Latorella 2002).

The level of human intervention within automation can be broadly divided into two major classes, Human Computer Interaction (HCI) and Human Agent Interaction (HAI), a subset of HCI itself (Norman 1998). In contrast to HCI, HAI agents permit more direct human intervention in their actions and dynamic adjustment of their parameters. The problem with the MITL approach is the need for rigorous monitoring and maintenance of each and every agent parameter and perhaps the manual computation of those parameters. Such a supervisor will be inundated with tedious decision making, which might jeopardise opportunities for more impactful, high-level decisions.

Man on the loop aims to herald a significant paradigm shift in controlling complex systems. Although there might be profound underlying phenomena, our simulation results will merely serve to show the path for prototypical system controls and not as validated scientific tautologies. Rigorous subsequent studies in collaboration with social scientists may yield enduring theories beyond the scope of our present endeavour. By and large, we are providing tools for the designers of larger systems to parametrically adjust their system for desired performance and characteristics. The present work provides articulation of a reasonable platform for significant exploration and development in this area.

In Section 2, we outline motivations for MOTL including the OODA loop. Section 3 offers basic behaviour articulation parameters developed by MOTL. Section 4 provides details of the implementation of these parameters followed by validation and observations in Section 5. Concluding remarks on MOTL are given in Section 6.

2. Guiding inspiration for MOTL

The OODA Loop, i.e. Observation, Orientation, Decide, Act (Figure 1); has been among the most popular methods for modelling decision-making in adversarial conditions. An entity must *observe* (O) her opponent's actions, *orient* (O) herself to the unfolding situation, *decide* (D) on the most appropriate course of action, and then *act* (A) (Luzwick 2000). In particular, the OODA loop was designed to allow World War II fighter pilots to get inside their opponent's decision cycle during a dogfight in an attempt to predict actions and responses to actions, thereby paralysing the enemy by operating inside the opponent's

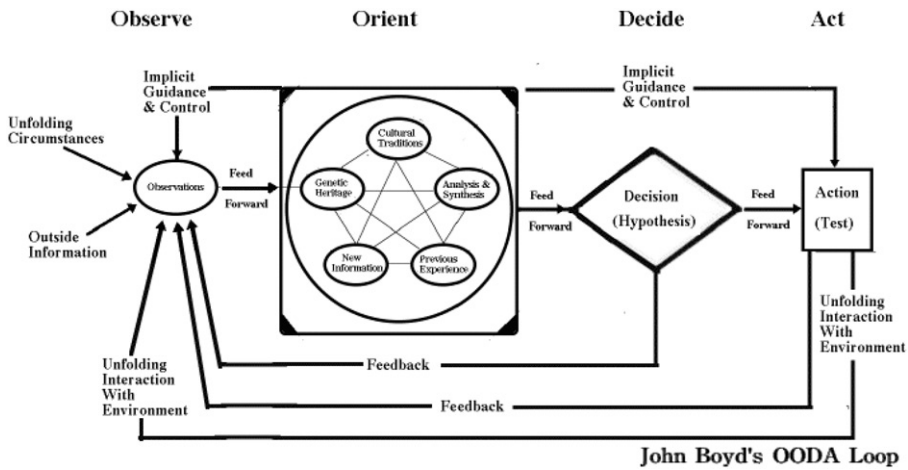


Figure 1. The OODA Loop. Adapted from (Coram, 2002).

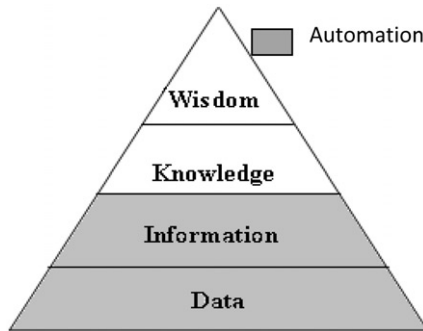


Figure 2. Different levels of human supervision.

OODA loop (Coram 2002). In the OODA Loop, one must continually observe, orient, decide and act, intertwining current decisions with future ones. One does so through a combination of rapidity, variety, harmony and initiative.

Any information processing method encompasses several refinement levels (Rich and Knight 1990). These layers could be easily presented using a pyramid (Figure 2). Consider intelligence analysts that on average spend 90% of their time moving, processing, and searching for essential elements of information (EEI) rather than their primary task of analysing, deciding and acting (Clark 2006). Even a small platform for performing intelligence, surveillance and reconnaissance (ISR) requires extensive juggling and management of cognitive tasks that are beyond the capabilities of a single person. Hence, to increase the efficiency of such analysts, we need to automate the lower layers in this pyramid and shift the level of human supervision to the upper layers (e.g. knowledge and wisdom).

Automating the tedious lower levels will allow the supervisor to focus on performing high-level cognitive tasks more efficiently. The supervisor's role with regards to such a system is like that of a parent to child. The supervisor allows the child to do the acting

while monitoring the child's behaviour and providing explicit feedback and direction. Such automation will positively affect capacity, safety, and efficiency of the whole system and will reduce personnel and maintenance costs (Sheridan 2002). Although our approach does not automate the decision-making processes, it provides a guideline for a more natural role for the human supervisor, i.e. MOTL.

3. Man on the loop

This section introduces key attributes that characterise the core components of the MOTL paradigm. Generally, these attributes are based on well-founded theories in social sciences and fall into three categories of personality, culture, and social reasoning. These attributes have long been implicitly considered to be important in intelligent human-machine interactions and mixed initiation. Our exposition is to make them explicit and propel them into the realm of supervisory control. Aiming for an intuitive understanding, we attempt to provide the reader with a sufficient social science background. However, a survey is outside our scope. Later in this article, these MOTL attributes are used to form parameters for a MOTL operator to adjust supervisory influence over the system.

In the remainder of this section, we briefly outline these attributes along with their significances for MOTL.

3.1. Personality

Personality and emotions are the most important aspects of generating believable social behaviours (Rousseau and Hayes-Roth 1998). *Personality* is the set of psychological traits that distinguish one individual from others. These traits correspond to patterns of behaviour and modes of thinking that determine a person's adjustment to the environment. In most psychological personality theories, agent behaviours are shown to be goal-driven, and action selection is based on personality, emotions, and relationships with others.

To parameterise MOTL based on personality, we define four characteristics: conformity, sociability, commitment, and disposition. Similar to (Hexmoor, McLaughlan and Baker 2005) where these characteristics were employed to model behaviour of a swarm of unmanned air vehicles, conformity and sociability in MOTL will reflect an agent's behaviour with respect to other working entities (operator or agent), while commitment and disposition determine an agent's dedication to achieving goals.

Conformity determines how quickly and exactly an agent responds to suggestions from a superior agent or the human operator. Some degree of conformity is always essential if individuals are to coexist peacefully, as when agents assume their rightful place in a waiting queue (Brehm, Kassin and Fein 2001). Social values are employed here to avoid conflicts that may appear among these autonomous and self-interested agents (Lopez and Luck 2003). Such conflicts are deemed unacceptable through definitions of social values which in turn define responsibilities and authorities for each agent and establish regularities of behaviour (Liu, Sun, Dix and Narasipuram 2001).

Sociability describes how well an agent obeys social values of the community, or in simpler terms, how gregarious the agent will be. The sociability of the individual in a group is crucial to the capability of the group as a whole. For example, if agents develop

sociability, they can interact more effectively and efficiently with one another in their community (Hanming and HuaCan 2004).

Socially responsible agents retain their local autonomy but still draw from and supply resources to the larger community. Through empirical evaluation, it has been shown that such agents produce both good system-wide performance and satisfactory individual performance (Jennings and Campos 1997). Therefore, conformity delineates sincerity towards superiors where as sociability outlines agents behaviour within its peer group.

In MOTL, commitment and disposition outline the agent's goal achievement qualities while working in a cooperative environment (Castelfranchi 1998). *Commitment* can be expressed as a percentage measure of the extent to which a goal is pursued (Barber and Han 1998). Commitment can also be used to determine if goals are automatically adopted when a role is assumed and the ways in which they can be discharged or cancelled (Cavedon and Sonenberg 1998). *Disposition* expressed in time is a measure of temperament, which sets the rate of frustration among agents for target acquisition, formally also known as *neuroticism*.

3.2. Culture

Formally, *culture* is defined as the shared patterns of behaviours and interactions, cognitive constructs, and affective understanding that are learned through a process of socialisation. In MOTL, culture is coupled not only to an agent's commitment to goal achievement and its communication style, but also to overt rules for operating errors and to an agent's motivation, morale, and perception of errors. Moreover, to avoid cultural differences which may tend to decrease coordination and understanding in an agent community, we have considered vital cultural parameters given by Hofstede (2001). These parameters are power distance index, individualism, uncertainty avoidance index, and long-term orientation. In MOTL terms, the terms "society" and "country" in Hofstede's research are replaced by "group" and "community", respectively. As an example, automation is typically correlated to high-power distance index and uncertainty avoidance index while maintaining a low level of individualism.

Power distance index (PDI) refers to the extent to which less powerful members expect and accept unequal power distribution within the culture. This index has been incorporated into MOTL to accommodate inherited or developed inter- and intra-power differences among several working agent communities in a complex system. PDI focuses on the degree of equality or social inequality among agents in the community and measures how much a culture has respect for authority. A high PDI ranking indicates the inequalities of power and social values like autonomy, trust, reliance, and benevolence will be allowed to grow within a group. In such a system, agents may view the supervisor as a benevolent dictator and obey her orders. Conversely, a low PDI ranking indicates that the group deemphasises the differences between agent's power and social values. Hence, equality and opportunity for every agent is stressed in this case.

To encompass every possible agent and its relations to cultures, MOTL includes two basic cultural orientations which are complementary to each other. Some cultures value individualism and the virtues of independence, autonomy, and self-reliance, while other value collectivism and the virtues of interdependence, cooperation, and social harmony (Brehm et al. 2001). *Individualism* (IDV) is the measure of the attitude that personal goals take priority over group allegiances. A high individualism ranking indicates that

individuality and individual rights are paramount within the group, where as a low individualism ranking typifies groups of a more collectivist nature with close ties between agents.

The *uncertainty avoidance index* (UAI) focuses on the level of tolerance for uncertainty and ambiguity within the group. A culture with a low UAI places higher value on risk-taking behaviours such as the exploration of new solutions that could potentially fail. Conversely, a culture with a higher UAI would tend to rely upon established but potentially inefficient solutions and would create a rule-oriented society that institutes laws, rules, regulations, and controls in order to reduce the amount of uncertainty.

Long-term orientation (LTO) focuses on the degree to which the society embraces traditional, forward thinking values. A high long-term orientation ranking indicates the community prescribes to the values of long-term commitments and respect for traditions. By varying a culture's long term orientation, individuals can benefit from the establishment of long-term commitments between groups, such as forming interest groups, or they can profit from short-term commitments such as the fulfilment of a specific goal.

3.3. Social reasoning

Autonomy, benevolence, trust, grouping, capability, power, and interdependence are the social notions that are used to qualify interaction and awareness in communities (Hexmoor and Poli 2003). Since autonomy, benevolence and trust have been explored elsewhere (Chang, Dillon and Hussain 2006), here we limit ourselves to grouping, capability, power, and interdependence.

We emphasise the notion of grouping in MOTL since when individuals are of the same ethnicity, results have shown that the agents within the group are socially more attractive and trustworthy (Scholtz and Bahrami 2003). Additionally, grouping helps in the easy and efficient delegation of unapproachable or unsuccessfully attempted goals.

To make this goal delegation process more accurate, the MOTL paradigm also incorporates the concept of *capability*. Capability can be defined as the agent's potential to complete a particular task. It must be a part of autonomy in the sense that an agent can be autonomous with respect to a particular task or action only if that agent has the ability to complete that task or action.

Similarly, social *power* plays a crucial role in delegating goals and roles from one agent to another as power is an ability to resist change or to exert change in others (Hexmoor 2002). Power and autonomy have a crucial relationship. Where power is the experience of social influence, autonomy is the experience of limits to liberty. Thus, an agent who experiences low social power is in a position to experience higher autonomy. However, experiencing low power from others is necessary but not sufficient for high autonomy.

Interdependence between the agents is related to the concepts of trust and autonomy and forms the vital basis for the functionality and existence of multi-agent systems. The nature of this interplay can be determined by the type and degree of interdependence between agents. For example, agents may have the potential to help, hinder, facilitate or negate the actions of others. These dependencies are an inherent property of any complex system and affects how the system's resources are divided between and controlled by the various agents.

Several interesting effects may result due to the interaction of autonomy, trust, and interdependence. For instance, when agents depend upon each other for goal achievement, inter-agent trust will develop. However, this may lead to the case where this dependency diminishes autonomy. However, the lack of dependency does not necessarily yield autonomy, and the effects of this dependency will depend upon the particular agent's personality (Beavers and Hexmoor 2004).

The social reasoning aspect of MOTL helps create a setting where agents can interact with agents they trust before asking the human user for permission. This reduces the number of queries to the human controller or supervisor while increasing the number of interactions between the agents. However, in the absence of trust such as when agents know little about each other, each agent interacts with the user independently and the number of interactions is high (Battul and Hexmoor 2003). As time goes by, inter-agent interactions will increase as trust relations form, reducing the supervisor intervention load. This phenomenon can be observed in Section 5.

4. Implementation

The computer simulation of social phenomena is a promising field of research at the intersection of social, mathematical and computer sciences. The objective of agent-based social simulation systems is to explore the dynamics of social phenomena replicated in a society of agents. In particular, the notion of an extended computational agent implementing cognitive capabilities (Doran 1998) encourages proliferations of artificial societies (Gilbert and Conte 1995; Epstein and Axtell 1996). We have selected to build a generic simulator which unlike traditional domain-specific simulators such as Gaudiano, Bonabeau and Shargel (2003) and Vincent, Horling, Wagner and Lesser (1998) remains domain neutral and can be tuned for any domain where a supervisor is required to control large systems or a community. The simulator can be adapted based on domain specifics.

4.1. Agents and goals

Several comments can be made about the simulation which will be presented. Agents are represented as active entities and can be used in any model without restriction to a specific domain. All agents are homogeneous in type but have varying capabilities and MOTL parameters. The capability of the agent with respect to a goal is the potential it has to achieve the goal.

The simulation centres on the completion of goals. *Goals* are defined as a non-empty set of attributes that describe states of affairs in the world (Lopez and Luck 2003). Like agents, goals do not belong to any particular domain, do not match any specific behaviour, and do not have any special property. The goals are described as abstract tasks to be performed, and the success or failure of the system is determined by the number of goals completed.

Goal achievement rate is the required capacity for an agent to achieve a particular goal and varies for each goal. The simulation concentrates on the MOTL parameters and how these parameters modulate goal achievement with respect to the change in supervisory control over the community.

4.2. Supervisor control over the agent community

A *supervisor* in an agent community gives directives to the agents and has some amount of control over those agents. Agents are treated as subordinates even when their actions are automated (Sheridan 2002). In MOTL, the supervisor will be continuously receiving information from the agents.

According to Sheridan, five roles can be identified for the human supervisor (Figure 3). The first role of the supervisor is to *plan* what steps have to be taken before passing any orders to the agents. The second role is to *teach* the agents how to act. This includes showing agents how to adhere to social values, how to achieve goals, and similar activities. In the third role, the supervisor is outside the loop to *monitor* the actions of agents. The supervisor detects if there are any agents which are failing to achieve their goals and if there any agents abusing their autonomous power. The fourth role is to *intervene* in the actions of agents whenever necessary. In such cases, the supervisor will adjust the parameters for the agents to correct the problem. In the last role, *learn*, the supervisor evaluates the performance of the agents and gains knowledge from the experience.

The human supervisory model iterates constantly and has some requirements placed upon the agents that will interact with the supervisor. In particular, the agent should possess the capacity to determine whether or not it is receiving relevant information from the supervisor. In addition, it should understand what actions it needs to perform and when it is appropriate to provide feedback to the supervisor (Figure 4). This feedback can utilise the agent's increasing experience to make recommendations for future plans and allow the supervisor to perceive patterns that can streamline the process.

4.3. Algorithm and functions

The algorithm utilised by the simulation (Figure 5) requires the computation of several values. Random values ranging from 1 to 100 were assigned to *capability* values of each agent with respect to each goal and for the *goal satisfaction rate* of each goal. For simplicity, the benevolence value is equal for all agents and increases along with the supervisory control over the community. *Delta value* is a small constant used in the algorithm which is used to increase the capability value for agents.

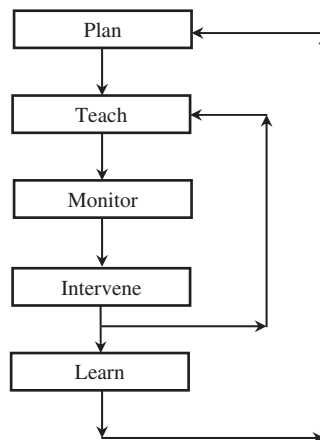


Figure 3. Human Supervisory Model. Adapted from (Sheridan 2002).

The *power* parameter is the basis on which the agents are divided into groups. Based on the input *number of groups*, the *margin* value is computed using the ratio $100/(\text{number of groups})$. Using this margin value, the *margin rate* for each group is computed. For example, if the number of groups is three, margin value will be $100/3 = 33.33$. So the three groups will be divided each with a margin of 1–33.33, 33.34–66.66, and 66.67–100. For each agent, if the power value falls in one of the margin values, the agent is sent into that group.

Various equations are utilised in the simulation to represent the interactions among agents and between agent and supervisor. These equations are essentially the quantisation of MOTL parameters. Although some parameters such as trust and autonomy can have many different definitions, the equations are mathematical representations of the descriptions provided in the previous section.

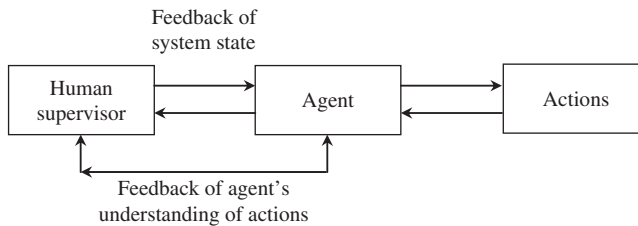


Figure 4. Agents feedback their command understanding.

- 1) Randomly allocates social power to every agent in the community.
- 2) Split up the agents into different groups based on their power.
- 3) Compute autonomy of each agent towards each goal in the community.
- 4) Calculate individual trust of each agent towards all other agents for every goal within the community.
- 5) Compute success or failure of goals based on the goal pursuing agent's capabilities.
- 6) Update the capability values of each goal pursuing agent, based on success or failure rate.
- 7) Recalculate the personality, culture, and social reasoning values using updated capability values, autonomy, and trust.
- 8) Repeat steps from (2) to (6) for a fixed number of cycles. Every time increase the benevolence and human supervision parameters.
- 9) Plot graphs using different human supervision values against the personality, culture, and social reasoning values.

Figure 5. High-level simulation control loop.

4.3.1. *Autonomy and overall trust*

Individual trust is computed using Equation (1)

$$\text{Individual trust } (B, A, G) = \text{Benevolence} + \text{Capability } (A, G) \quad (1)$$

where *Individual trust* (B, A, G) is the trust agent B has towards the agent A with respect to goal G and *Capability* (A, G) is the capability value of agent A for goal G .

Autonomy is computed using Equation (2)

$$\text{Autonomy } (A, G) = \sum_{i=1}^n \text{Individual trust } (A, B_i, G) + \text{Capability } (A, G) \quad (2)$$

where *Autonomy* (A, G) is the autonomy of agent A for the goal G and *Individual trust* (A, B_i, G) is the trust the agent A has on agent B_i for the goal G .

4.3.2. *Updating capability values*

For an agent to achieve a goal successfully, its capability value must be equal to or greater than the goal satisfaction rate for that particular goal:

$$\text{Capability } (A, G) \geq j \quad (3)$$

where j is goal satisfaction rate for goal G . If Equation (3) is satisfied, then agent A can satisfy goal G . If satisfied, the capability value for the goal by that agent is incremented by a delta value as an incentive. Otherwise, it is decreased by the delta value, and the goal is passed on to other agents within the same group.

4.3.3. *Goal delegating by one agent to another*

An agent A can delegate its goal G to another agent B if and only if:

- Both, agents A and B belong to the same community.
- Agent B 's capability value is highest in the community and is greater than or equal to the goal satisfaction rate for that goal G .
- The power of B is less than or equal to the power of A .

If none of the agents are able to fulfil the above three conditions in the community, the goal will remain unsatisfied.

An agent has failed to achieve a goal if its capability value is lower than goal satisfaction rate for that goal. Thus, failure rate is the rate at which the agent fails to achieve the goal. Failure rate is computed for each agent using the following formula,

$$\text{Failure Rate } (A) = \text{Failure Rate } (A) + (1/\text{total number of goals}) \quad (4)$$

where *Failure Rate* (A) is the failure rate for agent A and $1/\text{total number of goals}$ gives the failure rate for an agent to achieve one goal.

4.3.4. *Sociability*

An agent's autonomy is inversely proportional to sociability. As an agent with high autonomy will be highly independent, the probability of the agent adhering to social values will be low. Over time, the supervisor's control of the system enforces agents to follow

social values in the community. With increase in the supervisory control, the overall community sociability also increases apart from the individual agent's personal value possessed by an agent. This effect can be described as:

$$\text{Sociability } (A) = \text{Sociability } (A) + \text{Human Supervisor Control} \quad (5)$$

where *Sociability* (*A*) is the agent *A*'s personal sociability value.

4.3.5. Conformity

Condition 1: undelegated tasks

The conformity value for an agent is given by

$$\text{Conformity } (A) = \text{Conformity } (A) + (\text{Failure rate} * \text{Human Supervisor Control}) \quad (6)$$

where *Conformity* (*A*) is the conformity value for agent *A*. As the human supervisory control increases over the community, her effect on the community increases as well. As an agent continues to fail, the probability that the agent listens to the supervisor increases.

Condition 2: delegated tasks

After an unsuccessful attempt to satisfy a goal, if that goal is delegated to and satisfied by some other agent according to rules explained above, then the conformity value of the later agent will be increases by the power difference between the two agents:

$$\text{Conformity } (B) = \text{Conformity } (B) + (\text{Power } (A) - \text{Power } (B)) \quad (7)$$

4.3.6. Disposition

Condition 1: success

If the agent satisfies the goal assigned to it, the frustration level of agent is set to zero.

$$\text{Disposition } (A) = 0 \quad (8)$$

Condition 2: failure

If the agent *A* fails to satisfy the goal, then

$$\text{Disposition } (A) = \text{Disposition } (A) + \text{Frustration } (A, G) + \text{conformity } (A) \quad (9)$$

where *Disposition* (*A*) denotes agent *A*'s disposition value and

$$\text{Frustration } (A, G) = \text{Goal Satisfaction Value } (G) - \text{Capability } (A, G) \quad (10)$$

If agent *A* is unable to satisfy the goal, the disposition value is the summation of the frustration level and the help it receives from the supervisor (the value with which it listens to the supervisor suggestions) in achieving the goal *G*.

4.3.7. Commitment

Condition 1: success

If the agent satisfies the goal, then

$$\text{Commitment } (A) = \text{Commitment } (A) + \text{Human Supervisory Control} \quad (11)$$

where *Commitment (A)* is the commitment value of agent *A*. An agent becomes more dedicated toward achieving its goal if it has additional pressure imposed on it. Here, the pressure is simply the supervisory control over the agent.

Condition 2: delegation

If agent *A*'s goal is delegated to agent '*B*', then

$$\begin{aligned} \text{Commitment (B)} &= \text{Commitment (B)} + \text{Power Difference (A, B)} \\ &+ \text{Human Supervisory Control} \end{aligned} \quad (12)$$

Beyond the pressure imposed by the supervisory control, it is the additional power of agent *A* over agent *B* that forces agent *B* to become more dedicated to achieving the goal. An agent socially committed to another loses some of its autonomy. For example, a social commitment by agent *A* to agent *B* results in agent *B* having power and control over agent *A*. This leads agent *A* to adopt goals on behalf of agent *B*. However, as in human organisations, this form of commitment can be used as an input to the agent's decision-making and generate useful group discussions (Cavedon and Sonenberg 1998). Goal adoption is a basic ingredient of cooperative social behaviour (Castelfranchi 1998).

4.3.8. *Power distance index*

According to Hofstede, as the supervisory control increases, the PDI value increases. Therefore,

$$PDI(A) = PDI(A) + \text{Human Supervisory Control} \quad (13)$$

where *PDI (A)* is the power distance index of agent *A*.

4.3.9. *Long-term orientation*

As agents join the community, it is their long term commitment to stay in the community.

$$LTO(A) = LTO(A) + (\text{number of agents}/100) \quad (14)$$

Subsequently, an agent joins a group and develops a commitment to that group.

$$LTO(A) = LTO(A) + (\text{number of agents in that particular group}/100) \quad (15)$$

where *LTO (A)* is the long-term orientation for agent *A*.

4.3.10. *Uncertainty avoidance index*

As the supervisory control increases over the community, rules and regulations in the community becomes more rigid. Therefore, the uncertainty factor decreases as the supervisory control increases. Hence,

$$UAI(A) = UAI(A) + \text{Human Supervisory Control} \quad (16)$$

where *UAI (A)* is the uncertainty avoidance index for agent *A*.

4.3.11. *Individualism*

Individualism is the personal independence the agent has to act according to its own desires. Hence,

$$IDV(A) = IDV(A) + (100 - \text{Conformity}(A)) \quad (17)$$

where the *Conformity* value is the amount with which it obeys the supervisor.

If agent *A* delegates the goal to agent *B* then,

$$IDV(B) = IDV(B) - Power\ Difference(A, B) \quad (18)$$

where *Power Difference* (*A*, *B*) = *Power* (*A*) – *Power* (*B*). As agent *B* has to obey agent *A* in achieving a goal, agent *B* diminishes its individualism by not choosing its own desire for achieving goals. The reason behind agent *B* obeying agent *A* is that agent *A*'s power is greater than agent *B*. Thus, agent *B* loses its individualism by the differences of power between both agents *A* and *B*.

4.3.12. Interdependency

When agent *A* delegates a goal to agent *B*, then

$$Interdependency = Capability(A, G) - Capability(B, G) \quad (19)$$

Since agent *A*'s capability is not enough to achieve goal *G*, it will depend on agent *B* for achieving that goal. The value which made agent *A* dependant upon the agent *B* is their difference in capability values. Therefore, the amount with which agent *A* depends on agent *B* is the difference between the agent's capability values. Additionally,

$$Autonomy(B, G) = Autonomy(B, G) - PowerDifference(A, B) \quad (20)$$

$$Autonomy(A, G) = Autonomy(A, G) + Power\ Difference(A, B) \quad (21)$$

The reason for agent *B* for being dictated by agent *A* to achieve the goal *G* is that agent *A*'s power is greater than agent *B*'s power. Hence the difference of their powers is the value with which the agent *B* will lose its autonomy and agent *A* will gain autonomy.

$$Trust(A, B, G) = Trust(A, B, G) + Interdependency \quad (22)$$

Since the goal *G* of agent *A* is achieved by agent *B*, the trust of agent *A* increases towards agent *B* with respect to goal *G* because agent *B* helped agent *A* in achieving its goal. The amount of increase in the trust is the amount with which agent *A* depends on agent *B*, which is the same as the interdependency value.

5. Experimental results and observations

The results shown in this section were obtained by averaging multiple simulation runs that met the statistical significance criteria. Typical runs included 10 agents pursuing 20 goals while simultaneously varying MOTL parameters. Figure 6 is produced when the supervisor control is linearly increased over the community of the agents. This figure shows that supervisory control has an inverse parabolic effect on the MOTL parameters over the agent community. This is because all parameters except interdependence showed positive effects as the supervisory control goes on increase. This effect is summarised in the following linear combination:

$$\begin{aligned} MOTL\ Parameters = & IDV + PDI + LTO + UAI + Sociability + Commitment \\ & + Conformity + Disposition + Autonomy - Interdependence \end{aligned} \quad (23)$$

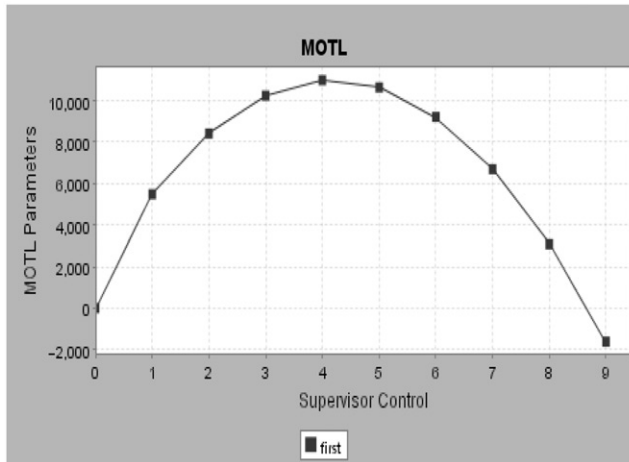


Figure 6. Effect of increase in supervisory control over MOTL parameters.

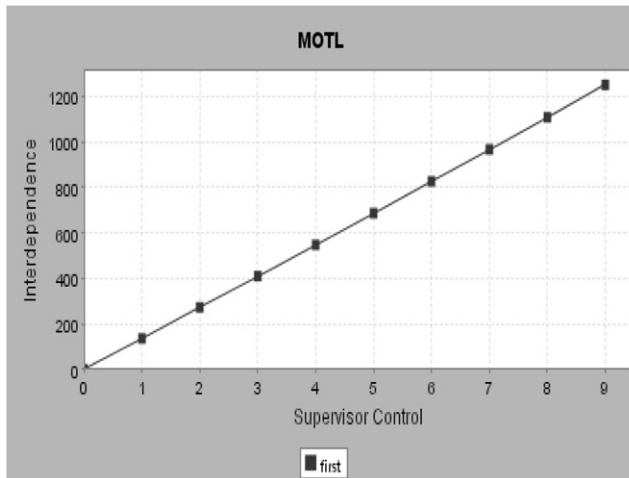


Figure 7. A typical plot of interdependence level vs. supervisory control.

This result suggests that there is a level of human intervention that will produce optimal system performance. This is inspired and akin to the level of arousal that yields optimal human performance established by the Yerkes–Dodson scientific principle (Yerkes and Dodson 1908). Thus, the experiment shows that the simulation follows real-world patterns.

Additional information can be gleaned from this experiment. Figure 7 shows that interdependency among agents increases with an increase in supervisory control. This in turn increases trust between individuals. This at some point may dominate the supervisory control, and agents may start rejecting commands. If the interdependence totally dominates the supervisory control, agents will start responding in a negative way with the supervisor. Therefore, as the supervisory control increases over an agent community,

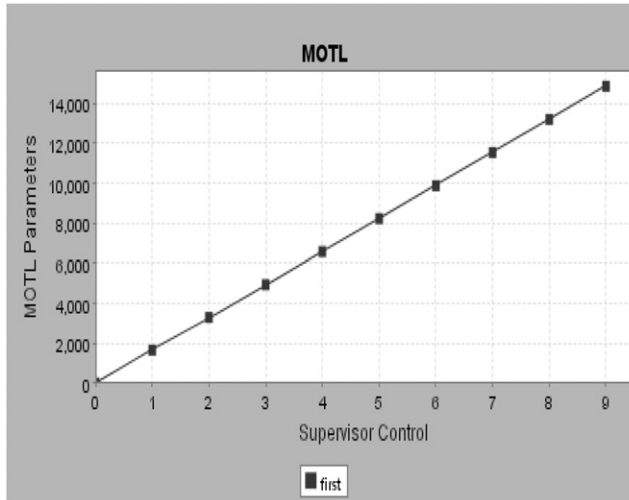


Figure 8. Improved supervisory effect over MOTL parameters.

it initially shows positive results but eventually will have a negative effect on the MOTL parameters, reducing the overall performance of the agent community.

Following this experiment, additional modifications were performed to evaluate the influence of specific variables. When individuality, power distance index, and autonomy are eliminated from the MOTL parameters:

$$\begin{aligned} \text{MOTL Parameters} = & LTO + UAI + \text{Sociability} + \text{Commitment} \\ & + \text{Conformity} + \text{Disposition} \end{aligned} \quad (24)$$

supervisory control has a linear effect on the MOTL parameters (Figure 8). Thus, limiting the effect of these three parameters will result in the linear improvement in the efficiency of the agent community.

If autonomy and interdependence are discarded:

$$\begin{aligned} \text{MOTL Parameters} = & IDV + PDI + LTO + UAI + \text{Sociability} \\ & + \text{Commitment} + \text{Conformity} + \text{Disposition}. \end{aligned} \quad (25)$$

the supervisory control still shows the negative effect but at slower rate (Figure 9). This is because these parameters suppress supervisory power over the agent community.

Additionally, this experiment explored the effect of autonomy of an agent versus the performance of supervisor control. When everything is moving smoothly, each agent in the system will linearly increase with same proportion to supervisory control (Figure 10). However, supervisor control has a decreasing non-linear effect (Figure 11) on the MOTL parameters if agents use their autonomy against the supervisory control but not for the overall benefit of the community. Hence, the performance of the community will get deteriorate as per following relation:

$$\begin{aligned} \text{MOTL Parameters} = & IDV + PDI + LTO + UAI + \text{Sociability} + \text{Commitment} \\ & + \text{Conformity} + \text{Disposition} - \text{Autonomy} - \text{Interdependence}. \end{aligned} \quad (26)$$

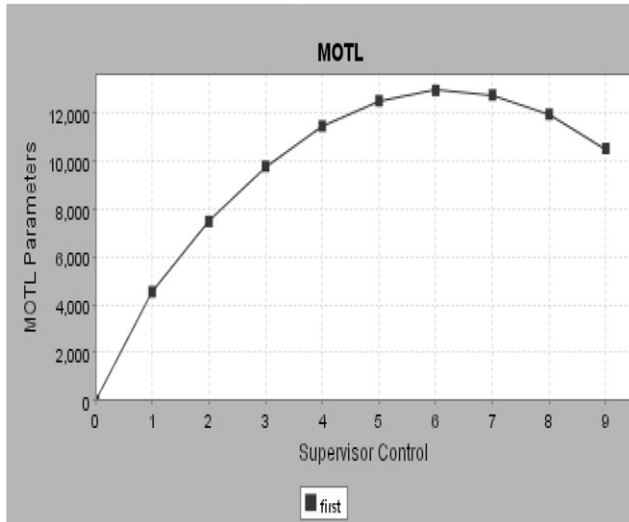


Figure 9. A typical plot of supervisory control vs. MOTL parameter (discarding autonomy and interdependence).

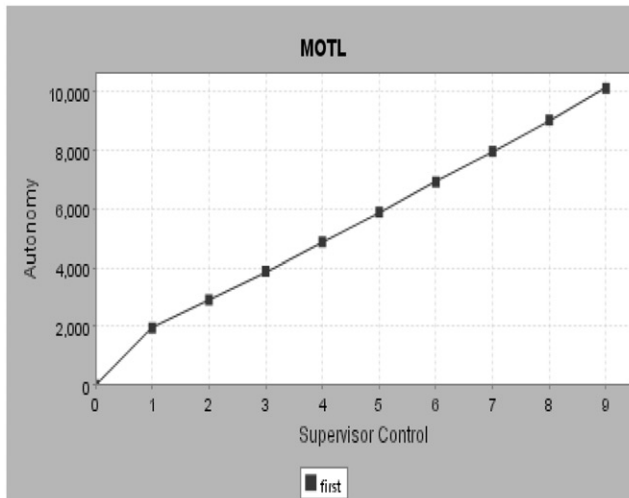


Figure 10. Supervisory control vs. autonomy with positive agent response to supervisor.

6. Conclusions

Real-time multi-agent systems built upon ontological knowledge backbones are rapidly emerging as viable platforms for implementing many applications, including net-centric warfare. The Enhanced ELINT Processor (EEP) is one such multi-agent system that automatically manages and controls ELINT sensor systems. During the early development of EEP, it was quickly recognised that new paradigms for monitoring and influencing automated systems were needed. Unlike current ELINT collection systems, automated systems cannot pause to attain human user review data prior to making decisions.

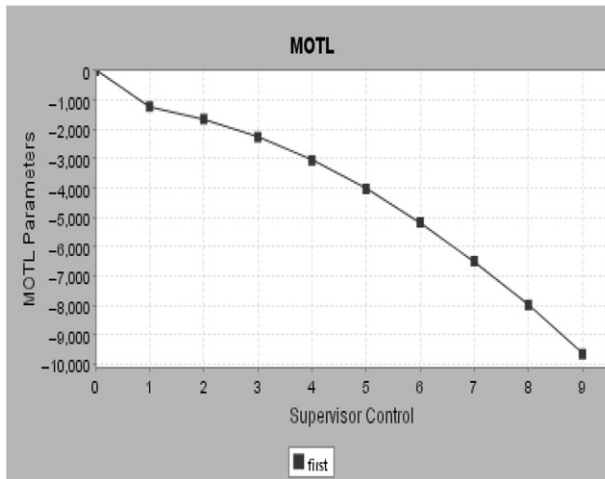


Figure 11. Deteriorated performance due to use negative use of autonomy on supervisor.

However, these highly complex systems still require capabilities for human supervision and influence.

The motivation for this work was to propose a major paradigm shift in the human supervision of complex systems. This article suggested a shift from the traditional Man in the Loop paradigm to a high-level control arrangement we termed MOTL. We conducted numerous experiments with an agent-based simulation to explore the effects of human supervision on an agent community. Observations of results indicate that the model used to simulate this paradigm follows the pattern known to exist in established, real-world systems where there exists an optimal level of control by the supervisor.

These positive results pave the way for further experimentation and expansion. In particular, plans are in place wherein the organisation of the agent system is incorporated into the MOTL paradigm. This will allow the system to take on characteristics of organisations that would allow it to better conform to the characteristics desired by the supervisor. Additionally, exploration of very large systems of agents could yield expanded knowledge of the effects of the MOTL paradigm shift on the efficient control of system behaviour.

Acknowledgements

This research is sponsored by the Air Force Research Laboratory (AFRL) under contract FA8750-06-C-0138. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of AFRL or the United States Government. Special thanks go to the Net Centric SIGINT Focused Information Enterprise (NCSFIE) at AFLR for its support.

References

- Barber, K.S., and Han, D.C. (1998), 'Multiagent Planning Under Dynamic Adaptive Autonomy', *IEEE Proceedings of the International Conference on Systems, Man, and Cybernetics*, 399–404.

- Battula, S., and Hexmoor, H. (2003), 'Towards Collaboration Between Human and Social Agents that Mind Human Social Personality', *Proceedings of the Fourth International Symposium on Collaborative Technologies and System*, 35, 205–210.
- Beavers, G., and Hexmoor, H. (2004), 'Types and Limits of Agent Autonomy', *Computational Autonomy, Potential, Risks, Solutions*, 2969, 95–102.
- Brehm, S.S., Kassin, S.M., and Fein, S. (2001), *Social Psychology* (5th ed.), Boston, MA, USA: Houghton Mifflin Company.
- Castelfranchi, C. (1998), 'Modeling Social Action of AI Agents', *Artificial Intelligence*, 103, 157–182.
- Castelfranchi, C., Falcone, R., and Hexmoor, H. (2003), 'A Prospectus of Agent Autonomy', *Agent Autonomy*, Boston, MA, USA: Kluwer, pp. 1–8. Chapter. 1.
- Cavedon, L., and Sonenberg, L. (1998), 'On social commitment, roles and preferred goals', in *Proceedings of the 3rd International Conference on Multi Agent Systems*, pp. 80–87.
- Chang, E., Dillon, T., and Hussain, F.K. (2006), *Trust and Reputation for Service-Oriented Environments*, New York, NY, USA: Wiley.
- Clark, R. (2006), *Intelligence Analysis: A Target-Centric Approach*, California, USA: CQ Press.
- Coram, R. (2002), *Boyd: The Fighter Pilot Who Changed the Art of War*, London: Little Brown and Company.
- Doran, J. (1998), 'Simulating Collective Misbelieve', *Journal of Artificial Societies and Social Simulation*, 1, <http://www.soc.survey.ac.uk/JASSS/1/1/3.html>
- Epstein, J.M., and Axtell, R.L. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge: The Brookings Institution Press; Washington, DC: MIT Press.
- Gaudio, P., Bonabeau, E., and Shargel, B. (2003), 'Swarm intelligence: a new C2 paradigm with an application to control of swarms of UAVs', in *Proceedings of 8th ICCRTS Command and Control Research and Technology Symposium*, pp. 1–8.
- Gilbert, N., and Conte, R. (1995), *Artificial Societies: The Computer Simulation of Social Life*, London: UCL Press.
- Hanming, H., and HuaCan, H. (2004), 'Role-based social mental states of agents', in *Proceedings of 2nd International Conference on Autonomous Robots and Agents*, pp. 267–270.
- Hexmoor, H. (2002), 'A Model of Absolute Autonomy and Power: Toward Group Effects', *Journal of Connection Science*, 14, 323–333.
- Hexmoor, H., McLaughlan B., and Baker, M. (2005), 'Swarm control in unmanned aerial vehicles', in *Proceedings of International Conference on Artificial Intelligence (IC-AI)*, pp. 911–917.
- Hexmoor, H., and Poli, P. (2003), 'Effects of Reciprocal Social Exchanges on Trust and Autonomy', in *Proceedings of the 3rd International Conference on Intelligent Systems Design and Application*, pp. 159–170.
- Hofstede, G. (2001), *Culture's Consequences*, California, USA: Sage.
- Jennings, N.R., and Campos, J.R. (1997), 'Towards a Social Level Characterization of Socially Responsible Agents', *IEEE Transactions on Software Engineering*, 144, 11–25.
- Liu, K., Sun, L., Dix, A., and Narasipuram, M. (2001), 'Norm Based Agency for Designing Collaborative Information Systems', *Journal of Information Systems*, 11, 229–247.
- Lopez, F.L., and Luck, M. (2003), 'Modeling Norms for Autonomous Agents,' in *Proceedings of the 4th Mexican International Conference on Computer Science*, pp. 238–245.
- Luzwick, P.G. (2000), 'Situational Awareness and OODA Loops, Coherent Knowledge-Based Operations Applied', *Journal of Computer Fraud and Security*, 2000, 15–17.
- McFarlane, D.C., and Latorella, K.A. (2002), 'The Scope and Importance of Human Interruption in Human-Computer Interaction Design', *Journal of Human-Computer Interaction*, 17, 1–61.
- Norman, D. (1998), *The Psychology of Everyday things*, New York: Basic Books.
- Rich, E., and Knight, K. (1990), *Artificial Intelligence*, New York: McGraw-Hill.
- Rousseau, D., and Hayes-Roth, B. (1998), 'A Social-Psychological Model for Synthetic Actors,' in *Proceedings of International Conference on Autonomous Agents, Minneapolis, Minnesota*, pp. 165–172.

- Scholtz, J., and Bahrami, S. (2003), 'Human-Robot Interaction: Development of an Evaluation Methodology for The Bystander Role of Interaction', *IEEE Proceedings of International Conference on Systems, Man, and Cybernetics*, 4, 3212–3217.
- Sheridan, T.B. (2002), *Humans and Automations: System Design and Research Issues*, New York, NY, USA: Wiley.
- Strassner, J.C. (2003), *Policy-Based Network Management: Solution for the Next Generation*, San Francisco: Morgan Kaufmann.
- Vincent, R., Horling, B., Wagner, T., and Lesser, V. (2001), 'Implementing Soft Real-time Agent Control', *Proceedings of the Fifth International Conference on Autonomous Agents*, Montreal, Quebec, Canada, 355–362.
- Yerkes, R.M., and Dodson, J.D. (1908), 'The Relation of Strength of Stimulus to Rapidity of Habit-Formation', *Journal of Comparative Neurology and Psychology*, 18, 459–482.