

# TOWARDS TASK ALLOCATION DECISION SUPPORT BY MEANS OF COGNITIVE MODELING OF TRUST (EXTENDED ABSTRACT)\*

Peter-Paul van Maanen <sup>a,b</sup>      Kees van Dongen <sup>a</sup>

<sup>a</sup> *Department of Human in Command, TNO Defense, Security and Safety  
P.O. Box 23, 3769 ZG Soesterberg, The Netherlands*

*Email: {peter-paul.vanmaanen, kees.vandongen}@tno.nl*

<sup>b</sup> *Department of Artificial Intelligence, Vrije Universiteit Amsterdam  
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands  
URL: <http://www.few.vu.nl/~pp/AA>*

## 1 Introduction

An important issue in research on human-machine cooperation concerns how tasks should be dynamically allocated within a human-machine team in order to improve team performance. The ability to support humans in task allocation decision making requires a thorough understanding of its underlying cognitive processes, and that of relative trust more specifically. This paper presents a computational agent-based model of these cognitive processes and proposes an experiment design that can be used to validate theoretical aspects of this model.

## 2 Cognitive theory and modeling

Many theories in the human factors literature about the cognitive processes underlying human-machine task allocation decisions include a notion of relative trust, i.e. differences of trust in two agents. This notion is often represented by means of a *preference state*:

$$P_i(\tau_o, t) = T_i(j, \tau_o, t) - T_i(i, \tau_o, t) \quad (1)$$

where  $i, j \in Agents = \{H, M, *\}$ ,  $\tau_o \in Tasks$ , and agent  $*$  represents the infallible agent. Here  $T_i(j, \tau_o, t)$  is called a *trust state* and represents that agent  $i$  trusts agent  $j$  with respect to its performance in executing task  $\tau_o$  at time point  $t$  (see Equation 2 below).

The task allocation decision is also bounded by an *allocation preference threshold*, indicating when relative trust does not result in a preference state high or low enough to rely on an agent. In the context of the Equation 1 an agent  $i$  prefers allocation of  $\tau_o$  to  $j$  if and only if  $1 \geq P_i(\tau_o, t) > \theta_i(\tau_o, t)$  and to  $i$  if and only if  $-1 \leq P_i(\tau_o, t) < -\theta_i(\tau_o, t)$  at time point  $t$ . Here  $\theta_i$  represents the allocation preference threshold of agent  $i$ . The real interval  $[-\theta_i, \theta_i]$  describes the allocation indifference zone of the agent  $i$  which may change in time due to for instance costs of waiting [1].

The term *trust* is used to refer to a mental state, a belief of a cognitive agent  $i$  about the achievement of a desired goal through another agent  $j$  or through agent  $i$  itself [2]. Trust is considered to depend on the past experiences that coincide on discrete time points with events that affect the agent's trust state. Trust is dynamic, but it does not simply increase and decrease with positive and negative experiences. How trust changes by successes and failures depends on how increases and decreases in performance are interpreted and causally attributed. Experiences of agent  $i$  can be given by its evaluation of agent  $j$ 's performance by means of a comparison with

---

\*The full version of this paper appeared in: *Proceedings of the Eighth International Workshop on Trust in Agent Societies (Trust 2005)*, Utrecht, The Netherlands.

what agent  $i$  believes the infallible agent  $*$  would do if it was allocated to a task  $\tau_o$ :

$$T_i(j, \tau_o, t) = 1 - \frac{D_i(\sigma_i(j, \tau_o, t), \sigma_i(*, \tau_o, t))}{|\sigma_i(*, \tau_o, t)|} \quad (2)$$

where  $D_i$  is a function calculating the pre-normalized performance by means of distances between two strings, according to agent  $i$ . The function  $\sigma_i(j, \tau_o, t)$  returns a string of sequentially ordered actions resulting from the execution of task  $\tau_o$ , by agent  $j$ , according to agent  $i$ , and until time point  $t$ . Note that initial values should apply to trust states based on  $\sigma_i$ 's with lengths closer to 0. These values are typically more and more dependent on the inverse of the total number of possible actions.

### 3 Experiment design and validation

In order to validate implications of the above theory a simple experimental task is developed. The goal of this task is to predict, as a human-machine team, the location of a disturbance. In every trial a disturbance will occur at one of three locations. Each trial consists of three phases. First, human and machine predict the location of the next disturbance. Second, given both predictions, both human and machine decide which advised prediction should be selected. In the last phase the actual location of the disturbance is revealed and both human and machine can update their trust and preference states. In this experiment trust states will be manipulated by varying task difficulty, the length and complexity of the disturbances (TD), and machine reliability (MR). It is for instance expected that if preference states are in the indifference zone, errors are likely to occur (Figure 1).

TD × MR	TD1	TD2	TD3
100% MR	$-\theta_H \leq P_H \leq \theta_H$	$1 \geq P_H > \theta_H$	$1 \geq P_H > \theta_H$
70% MR	$-1 \leq P_H < -\theta_H$	$-\theta_H \leq P_H \leq \theta_H$	$1 \geq P_H > \theta_H$
50% MR	$-1 \leq P_H < -\theta_H$	$-1 \leq P_H < -\theta_H$	$-\theta_H \leq P_H \leq \theta_H$

Figure 1: The proposed 3 (task difficulty) × 3 (machine reliability) experiment design with the expected properties of corresponding allocation preference state  $P_H$ .

### 4 Discussion

After being confident on the replicability of previously found experimental findings in various domains in literature the experimental environment will be used for the further study of trust dynamics in MAS. This includes indirect acquisition of knowledge (e.g., reputation, gossip), analogical judgments, allocation engagement costs (e.g., waiting, cooperation, and overhead costs), allocation implementation errors, level of autonomy, allocation decision threshold, multi-tasking, task order, and time pressure.

### References

- [1] J. R. Busemeyer and A. Rapoport. Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32:91–143, 1988.
- [2] R. Falcone and C. Castelfranchi. Social trust: a cognitive approach. *Trust and deception in virtual societies*, pages 55–90, 2001.