Modeling Dynamics of Relative Trust of Competitive Information Agents

Mark Hoogendoorn, S. Waqar Jaffry, and Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence,
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
{mhoogen, swjaffry, treur}@few.vu.nl

Abstract. In order for personal assistant agents in an ambient intelligence context to provide good recommendations, or pro-actively support humans in task allocation, a good model of what the human prefers is essential. One aspect that can be considered to tailor this support to the preferences of humans is trust. This measurement of trust should incorporate the notion of relativeness since a personal assistant agent typically has a choice of advising substitutable options. In this paper such a model for relative trust is presented, whereby a number of parameters can be set that represent characteristics of a human.

1 Introduction

Nowadays, more and more ambient systems are being deployed to support humans in an effective way [1], [2] and [3]. An example of such an ambient system is a personal agent that monitors the behaviour of a human executing certain complex tasks, and gives dedicated support for this. Such support could include advising the use of a particular information source, system or agent to enable proper execution of the task, or even involving such a system or agent pro-actively. In order for these personal agents to be accepted and useful, the personal agent should be well aware of the habits and preferences of the human it is supporting. If a human for example dislikes using a particular system or agent, and there are several alternatives available that are more preferred, the personal agent would not be supporting effectively if it would advise, or even pro-actively initiate, the disliked option.

An aspect that plays a crucial role in giving such tailored advice is to represent the trust levels the human has for certain options. Knowing these trust values allows the personal assistant to reason about these levels, and give the best possible support that is in accordance with the habits and preferences of the human. Since there would be no problem in case there is only one way of supporting the human, the problem of selecting the right support method only occurs in case of substitutable options. Therefore, a notion of relative trust in these options seems more realistic than having a separate independent trust value for each of these options. For instance, if three systems or agents can contribute X, and two of them perform bad, whereas the third performs pretty bad as well, but somewhat better in than the others, your trust in that third option may still be a bit high since in the context of the other options it is the best alternative. The existing trust models do however not explicitly handle such relative trust notions [4] and [5].

This paper introduces an approach to model relative trust. In this model, a variety of different parameters can be set to fully tailor this trust model towards the human being supported. These aspects include initial trust and distrust, the weighing of positive and negative experiences, and the weight of past experiences. The model is represented by means of differential equations to also enable a formal analysis of the proposed model. Experiments have been conducted with a variety of settings to show what the influence of the various parameters is upon the trust levels.

This paper is organised as follows. First, in Section 2 the model is explained. Next, Section 3 presents a number of simulation results. In Section 4 the model is used to compare different cultures with each other. Section 5 presents a formal analysis of the model. Finally, Section 6 is a discussion.

2 Modelling Dynamics of Trust of Competitive Trustees

This section proposes a model that caters the dynamics of a human's trust on competitive trustees. In this model trust of the human on a trustee depends on the relative experiences with the trustee in comparison to the experiences from all of the competitive trustees. The model defines the total trust of the human as the difference between positive trust and negative trust (distrust) on the trustee. It includes personal human characteristics like trust decay, flexibility, and degree of autonomy (context-independence) of the trust. Figure 1 shows the dynamic relationships in the proposed model.

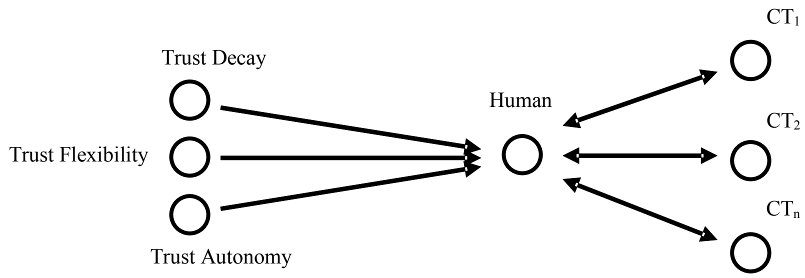


Fig. 1. Trust-based interaction with n competitive trustees (information agents IA)

In this model it is assumed that the human is bound to request one of the available competitive trustees at each time step. The probability of the human's decision to request one of the trustees $\{CT_1, CT_2, \ldots CT_n\}$ at time t is based on the trust value $\{T_1, T_2, \ldots T_n\}$ for each CT_i respectively at time t. In the response of the human's request CT_i gives experience value ($E_i(t)$) from the set {-1, 1} which means a negative and positive experience respectively. This experience is used to update the trust value for the next time point. Besides {-1, 1} the experience value can also be 0, indicating that CT_i gives no experience to the human at time point t.

2.1 Parameters Characterising Individual Differences between Humans

To tune the model to specific personal human characteristics a number of parameters are used.

Flexibility β The personality attribute called trust flexibility (β) is a number between [0, 1] that represents in how far the trust level at time point t will be adapted when human has a (positive or negative) experience with a trustee. If this factor is high then the human will give more weight to the experience at $t+\Delta t$ than the already available trust at t to determine the new trust level for $t+\Delta t$ and vice versa.

Trust Decay γ The human personality attribute called trust decay (γ) is a number between [0, 1] that represents the rate of trust decay of the human on the trustee when there is no experience. If this factor is high then the human will forget soon about past experiences with the trustee and vice versa.

Autonomy η The human personality attribute called autonomy (η) is a number between [0, 1] that indicates in how far trust is determined independent of trust in other options. If the number is high, trust is (almost) independent of other options.

Initial Trust The human personality attribute called initial trust indicates the level of trust assigned initially to a trustee.

2.2 Dynamical Models for Relative Trust and Distrust

The model is composed from two models: one for the positive trust, accumulating positive experiences, and one for negative trust, accumulating negative experiences. The approach of taking positive and negative trust separately at the same time to measure total trust is similar to the approaches taken in literature for degree of belief and disbelief [6] and [7]. Both negative and positive trusts are a number between [0, 1]. While human total trust at CT_i on any time point t is the difference of positive and negative trust at CT_i at time t.

Here first the positive trust is addressed. The human's relative positive trust of CT_i at time point t is based on a combination of two parts: the *autonomous* part, and the *context-dependent* part. For the latter part an important indicator is the human's relative positive trust of CT_i at time point t (denoted by $\tau_i^+(t)$): the ratio of the human's trust of CT_i to the average human's trust on all options at time point t. Similarly an indicator for the human's relative negative trust of CT_i at time point t (denoted by $\tau_i^-(t)$) is the ratio between human's negative trust of the option CT_i and the average human's negative trust on all options at time point t. These are calculated as follows:

$$\tau_i^+(t) = \frac{T_i^+(t)}{\sum_{j=1}^{n} T_j^+(t) / n} \quad \text{and} \quad \tau_i^-(t) = \frac{T_i^-(t)}{\sum_{j=1}^{n} T_j^-(t) / n}$$

Here the denominators $\sum_{j=1}^{n} T_j^+(t) / n$ and $\sum_{j=1}^{n} T_j^-(t) / n$ express the average positive and negative trust over all options at time point t respectively. The context-dependent part was designed in such a way that when the positive trust is above the average, then upon each positive experience it gets an extra increase, and when it is below average it gets a decrease. This models a form of competition between the different information agents. The principle used is a variant of a 'winner takes it all' principle, which for example is sometimes modelled by mutually inhibiting neurons representing the different options. This principle has been modelled by basing the change of trust upon a positive experience on $\tau_i^+(t) - 1$, which is positive when the positive trust is above average and negative when it is below average. To normalise, this is multiplied by a factor $T_i^+(t)*(1 - T_i^+(t))$. For the autonomous part the change upon a positive experience is modelled by $1 - T_i^+(t)$. As η indicates in how far the human is autonomous or context-dependent in trust attribution, a weighted sum is taken with weights η and $1-\eta$ respectively. Therefore, using the parameters defined in above $T_i^+(t+\Delta t)$ is calculated using the following equations. Note that here the competition mechanism is incorporated in a dynamical systems approach where the values of $\tau_i^+(t)$ have impact on the change of positive trust over time. Followings are the equations when $E_i(t)$ is 1, 0 and -1 respectively.

$$T_i^+(t+\Delta t) = T_i^+(t) + \beta * \left(\eta * \left(1 - T_i^+(t)\right) + (1-\eta) * \left(\tau_i^+(t) - 1\right) * T_i^+(t) * \left(1 - T_i^+(t)\right)\right) * \Delta t$$
$$T_i^+(t+\Delta t) = T_i^+(t) - \gamma * T_i^+(t) * \Delta t$$
$$T_i^+(t+\Delta t) = T_i^+(t)$$

Notice that here in the case of negative experience positive trust is kept constant to avoid doubling the effect over all trust calculation as negative experience is accommodated fully in the negative trust calculation. In one formula this is expressed by:

$$T_i^+(t+\Delta t) = T_i^+(t) + \begin{pmatrix} \beta * \left(\eta * \left(1 - T_i^+(t)\right) + (1-\eta) * \left(\tau_i^+(t) - 1\right) * T_i^+(t) * \left(1 - T_i^+(t)\right)\right) * E_i(t) \\ * \left(E_i(t)+1\right)/2 - \gamma * T_i^+(t) * \left(1 + E_i(t)\right) * \left(1 - E_i(t)\right) \end{pmatrix} * \Delta t$$

In differential equation form this can be reformulated as:

$$\frac{dT_i^+(t)}{dt} = \beta * \left(\eta * \left(1 - T_i^+(t)\right) + (1-\eta) * \left(\tau_i^+(t) - 1\right) * T_i^+(t) * \left(1 - T_i^+(t)\right)\right) * E_i(t) * \left(E_i(t)+1\right)/2$$
$$- \gamma * T_i^+(t) * \left(1 + E_i(t)\right) * \left(1 - E_i(t)\right)$$

Notice that this is a system of n coupled differential equations; the coupling is realised by $\tau_i^+(t)$ which includes the sum of the different trust values for all j. Similarly, for negative trust followings are the equations when $E_i(t)$ is -1, 0 and 1 respectively.

$$T_i^-(t+\Delta t) = T_i^-(t) + \beta * \left(\eta * \left(1 - T_i^-(t)\right) + (1-\eta) * \left(\tau_i^-(t) - 1\right) * T_i^-(t) * \left(1 - T_i^-(t)\right)\right) * \Delta t$$
$$T_i^-(t+\Delta t) = T_i^-(t) - \gamma * T_i^-(t) * \Delta t$$
$$T_i^-(t+\Delta t) = T_i^-(t)$$

In one formula this is expressed as:

$$T_i^-(t+\Delta t) = T_i^-(t) + \begin{pmatrix} \beta * \left(\eta * \left(1 - T_i^-(t)\right) + (1-\eta) * \left(\tau_i^-(t) - 1\right) * T_i^-(t) * \left(1 - T_i^-(t)\right)\right) * E_i(t) \\ * \left(E_i(t)-1\right)/2 - \gamma * T_i^-(t) * \left(1 + E_i(t)\right) * \left(1 - E_i(t)\right) \end{pmatrix} * \Delta t$$

In differential equation form this can be reformulated as:

$$\frac{dT_i^-(t)}{dt} = \beta * \left(\eta * \left(1 - T_i^-(t)\right) + (1-\eta) * \left(\tau_i^-(t) - 1\right) * T_i^-(t) * \left(1 - T_i^-(t)\right)\right) * E_i(t) * \left(E_i(t)-1\right)/2$$
$$- \gamma * T_i^-(t) * \left(1 + E_i(t)\right) * \left(1 - E_i(t)\right)$$

Notice that this again is a system of n coupled differential equations but not coupled to the system for the positive case described above.

2.3 Combining Positive and Negative Trust in Overall Relative Trust

The human's total trust $T_i(t)$ of CT_i at time point t is a number between [-1, 1] where -1 and 1 represent minimum and maximum values of the trust respectively. It is the difference of the human's positive and negative trust of CT_i at time point t:

$$T_i(t) = T_i^+(t) - T_i^-(t)$$

In particular, also the human's initial total trust of CT_i at time point 0 is $T_i(0)$ which is the difference of human's initial trust $T_i^+(0)$and distrust $T_i^-(0)$ in CT_i at time point 0.

2.4 Decision Model for Selection of a Trustee

As the human's total trust is a number in the interval [-1, 1], to calculate the *request probability* to request CT_i at time point t ($RP_i(t)$) the human's total trust $T_i(t)$ is first projected at the interval [0, 2] and then normalized as follows;

$$RP_i(t) = \frac{T_i(t)+1}{\sum_{j=1}^{n}\left(T_j(t)+1\right)}$$

3 Simulation Results

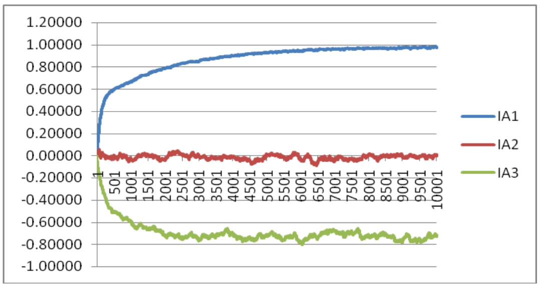
This section describes a case study to analyze the behavior of the model described in Section 2. This case study analyzes the dynamics of a human's total trust on the three competitive Information Agents (IA's). Several simulations were conducted in this case study. Few of the simulation results are presented in this and the next section. Other variations could be found in appendix A[1]. In this case study it is assumed that the human is bound to request one of the available competitive information agents at each time step. The probability of the human's decision to request one of the information agents $\{IA_1, IA_2, IA_3\}$ at time t is based on the human's total trust with each information agent respectively at time t $\{T_1(t), T_2(t), T_3(t)\}$ (i.e. the equation shown in Section 2.4). In response of the human's request for information the agent gives an experience value $E_i(t)$.

3.1 Relativeness

The first experiment described was conducted to observe the relativeness attribute of the model (see Figure 2). In the Figure, the x-axis represents time, whereas the y-axis represents the trust value for the various information providers. The configurations taken into the account are as shown in Table 1.

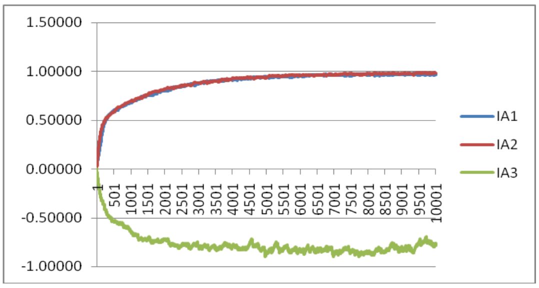
Table 1. Parameter values to analyze the dynamics of relative trust with the change in IAs responses.

Attribute	Symbol	Value
Trust Decay	γ	0.01
Autonomy	η	0.25
Flexibility	β	0.75
Time Step	Δt	0.10
Initial Trust and Distrust of $\{IA_1, IA_2, IA_3\}$	$T_1^+(0), T_2^+(0), T_3^+(0), T_1^-(0), T_2^-(0), T_3^-(0)$	0.50, 0.50, 0.50, 0.50, 0.50, 0.50

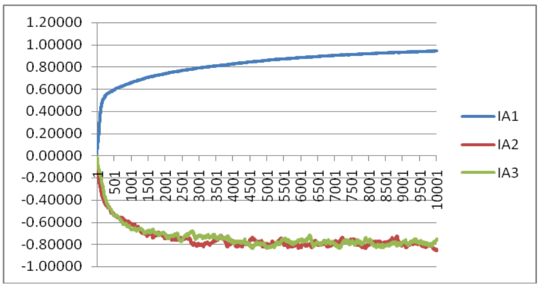


a)

[1] http://www.cs.vu.nl/~mhoogen/trust/appendix-CIA-2008.pdf



b)



c)

Fig. 2. Model Dependence on Amount of Positive Response from IAs: a) Information Agents IA_1, IA_2, IA_3 give experience positive, random (equal probability to give a positive or negative experience), negative respectively on each request by the Human respectively. b) Information Agents IA_1, IA_2, IA_3 give experience positive, positive, negative on each request by the Human respectively. c) Information Agents IA_1, IA_2, IA_3 give experience positive, negative, negative on each request by the Human respectively.

It is evident from above graphs that the information agent who gives more positive experience gets more relative trust than the others, which can be considered a basic property of trust dynamics (trust monotonicity) [5] and [8].

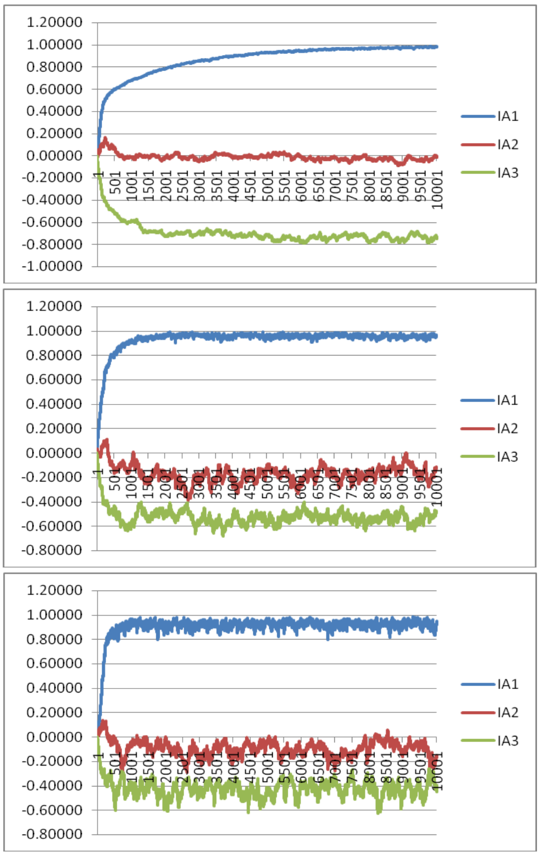
3.2 Trust Decay

This second experiment, shown in Figure 3, was configured to observe the change in the total trust in relation to change in the trust decay attribute γ of the human. The configurations taken into the account are as shown in Table 2.

Table 2. Parameter values to analyze the dynamics of relative trust with the change in trust decay (γ).

Attribute	Symbol	Value
Experience $\{IA_1, IA_2, IA_3\}$	E_1, E_2, E_3	1, random, -1
Autonomy	η	0.25
Flexibility	β	0.75
Time Step	Δt	0.10
Initial Trust and Distrust of $\{IA_1, IA_2, IA_3\}$	$T_1^+(0)$, $T_2^+(0)$, $T_3^+(0)$, $T_1^-(0)$, $T_2^-(0)$, $T_3^-(0)$	0.50,0.50,0.50, 0.50,0.50,0.50

In these cases also the information agent who gives more positive experience gets more relative trust than the others. Furthermore, if the trust decay is higher, then the trust value drops rapidly on no experience (see Figure 3c; more unsmooth fringes of the curve).



a)

b)

c)

Fig. 3. Model Dependence on Trust Decay: a) $\gamma = 0.01$. b) $\gamma = 0.05$. c) $\gamma = 0.10$.

3.3 Flexibility of Trust

This experiment is configured to observe the change in the total trust with the change in the human's flexibility of the trust (see Figure 4). Configurations taken into the account are shown in Table 3.

Table 3. Parameter values to analyze the dynamics of relative trust with the change in flexibility (β).

Attribute	Symbol	Value
Experience {IA_1, IA_2, IA_3}	E_1, E_2, E_3	1, random, -1
Trust Decay	γ	0.01
Autonomy	η	0.25
Time Step	Δt	0.10
Initial Trust and Distrust of {IA_1,IA_2,IA_3}	$T_1^+(0)$, $T_2^+(0)$, $T_3^+(0)$, $T_1^-(0)$, $T_2^-(0)$, $T_3^-(0)$	0.50, 0.50, 0.50, 0.50, 0.50, 0.50

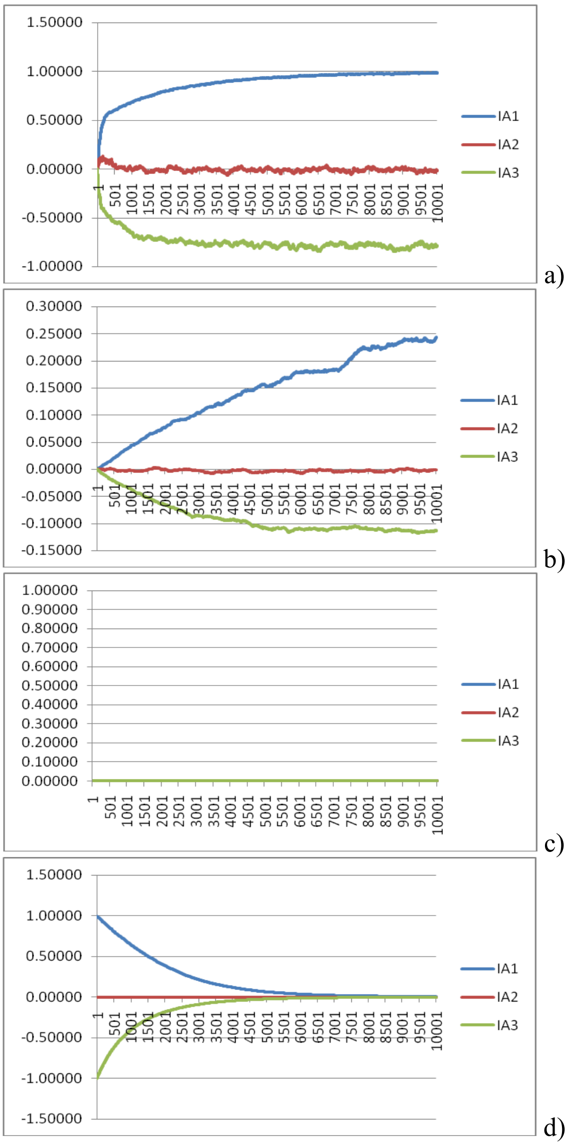


Fig. 4. Model Dependence on Trust Flexibility: a) $\beta = 1$, b) $\beta = 0.01$, c) $\beta = 0.00$, d) $\beta = 0.00$ and $T_1(0)=1$, $T_2(0)=0$, $T_3(0)=-1$.

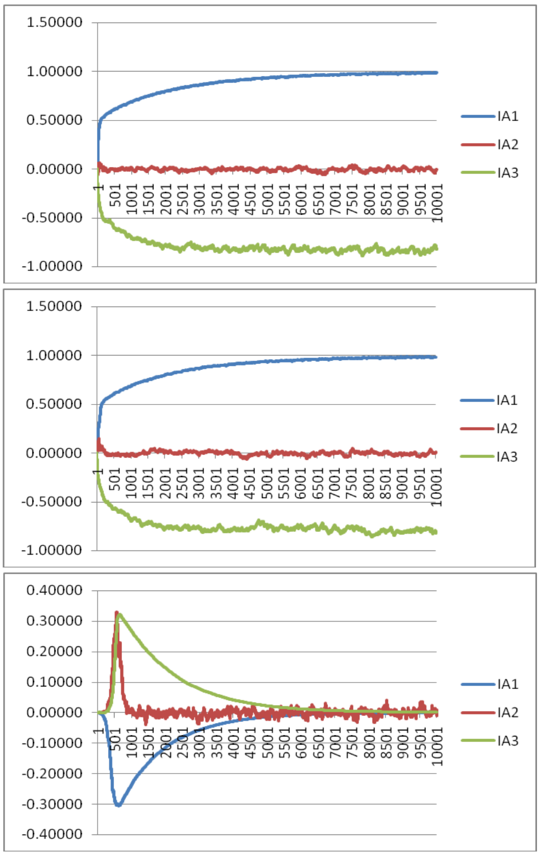
In these cases again the information agent who gives more positive experience gets more human's relative trust then the others. Furthermore as the values of the β decrease the rate of change of the trust also decrease. In Figure 4c, $\beta=0$ which means that trust does not change on experiences at all, so the initial values retain for experiences from the information agents hence trust value remains stable. Finally in the Figure 4d as initial values of the total trust are taken $T_1(0)=1$, $T_2(0)=0$ and $T_3(0)=-1$ instead of $T_1(0)=0$, $T_2(0)=0$ and $T_3(0)=0$, so the total trust decays due to the trust decay factor and becomes stable after a specific time span.

3.4 Autonomy of Trust

This experiment (see Figure 5) is configured to observe the change in the human trust with the change in the human's autonomy for the total trust calculation. Configurations taken into the account are shown in Table 4.

Table 4. Parameter values to analyze the dynamics of relative trust with the change in autonomy (η).

Attribute	Symbol	Value
Experience {IA_1, IA_2, IA_3}	E_1, E_2, E_3	1, random, -1
Trust Decay	γ	0.01
Flexibility	β	0.75
Time Step	Δt	0.10
Initial Trust and Distrust of {IA_1,IA_2,IA_3}	$T_1^+(0)$, $T_2^+(0)$, $T_3^+(0)$, $T_1^-(0)$, $T_2^-(0)$, $T_3^-(0)$	0.50, 0.50, 0.50, 0.50, 0.50, 0.50



a)

b)

c)

Fig. 5. Model Dependence on Trust Autonomy: a) η=1.0, b) η=0.50, c) η=0.00.

In these cases also the information agent who gives more positive experience gets more relative trust then the others. Further more as the values of the η decrease the human weights the relative part of the trust more than the autonomous trust. In Figure 5c, $\eta=0$ which means that the human does not take into account the autonomous trust. This gives unstable patterns that are extremely sensitive to the initial conditions of the system. The example graph shown is just one of these patterns.

3.5 Initial Trust and Distrust

This experiment is configured to observe the change in the total trust with the change in the human's initial trust and distrust ($T^{+}_{i}(0)$, $T^{-}_{i}(0)$) on information agents (see Figure 6). Configurations taken into the account are shown in Table 5.

Table 5. Parameter values to analyze the dynamics of relative trust with the change in initial trust.

Attribute	Symbol	Value
Experience {IA_1, IA_2, IA_3}	E_1, E_2, E_3	1, random, -1
Trust Decay	γ	0.01
Autonomy	η	0.25
Flexibility	β	0.75
Time Step	Δt	0.10

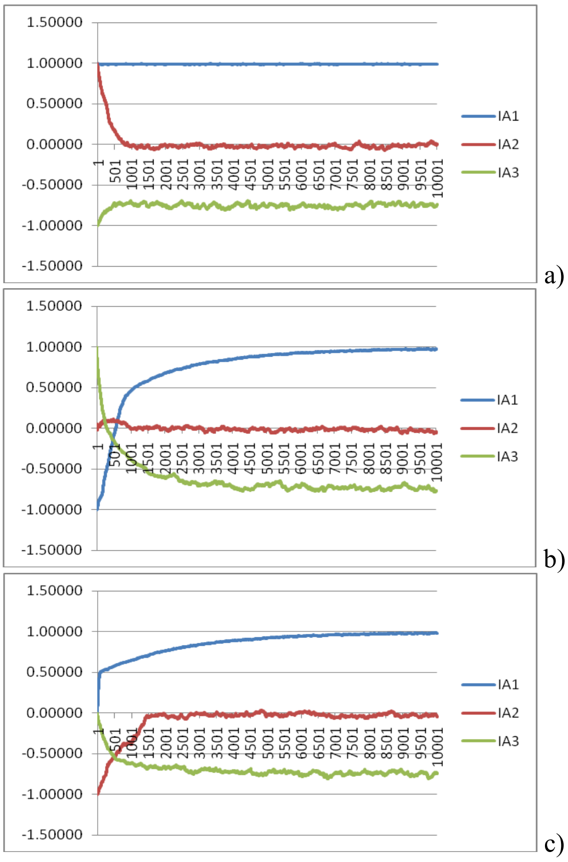


Fig. 6. Model Dependence on Initial Trust {$T_1(0)$, $T_2(0)$, $T_3(0)$}: a) 1, 1, -1. b) -1, 0, 1. c) 0, -1, 0.

It is observed from the above graphs that the final outcome of the trust is not very sensitive for the initial values.

4 Dynamics of Relative Trust in Different Cultures

The degree of reliability of available information sources may strongly differ in different types of societies or cultures. In some types of societies it may be exceptional when an information source provides 10% or more false information, whereas in other types of societies it is more or less normal that around 50% of the outcomes of information sources is false. If the positive experiences percentage given by the information agents varies significantly, then the total relative trust of the human on the these information agents may differ as well. This case study was designed to study dynamics of the human's trust on information agents in different cultures with respect to the percentages of the positive experiences they provide to the human. A main question is whether in a culture where most information sources are not very reliable, the trust in a given information source is higher than in a culture where the competitive information sources are more reliable. Cultures are named with respect to percentage of the positive experiences provided by the information agents to the human as shown in Table 6 and other experimental configurations in Table 7.

Table 6. Classification of Human Cultures with respect to the Positive Experiences given by the IAs.

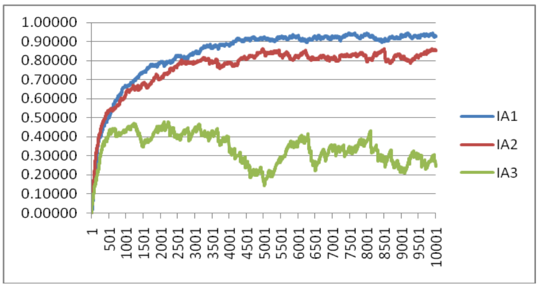
Culture Name	Percentage of the positive experiences by the information agents {IA_1, IA_2, IA_3}
A	100, 99, 95
B	50, 40, 30
C	10, 0, 0
D	0,0,0

Table 7. Parameter values to analyze the Relative Trust Dynamics in different Cultures.

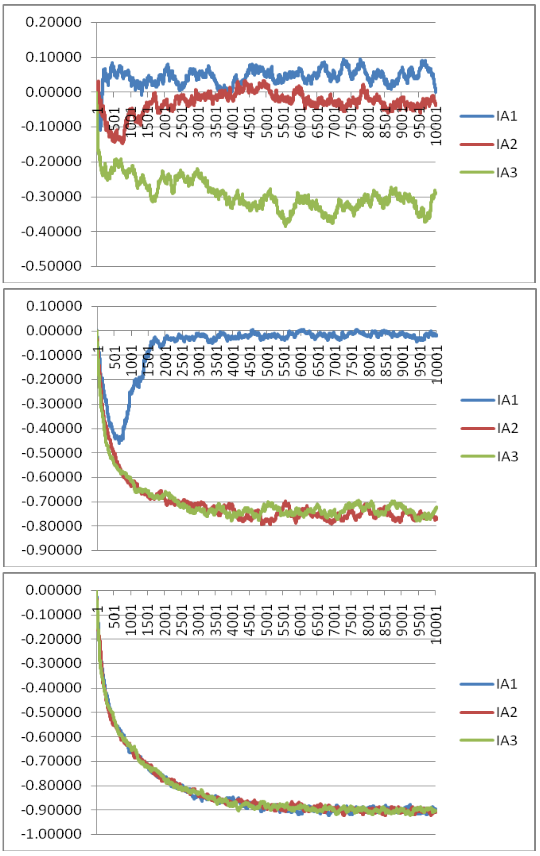
Attribute	Symbol	Value
Trust Decay	γ	0.01
Autonomy	η	0.25
Flexibility	β	0.75
Time Step	Δt	0.10
Initial Trust and Distrust of {IA_1,IA_2,IA_3}	$T_1^+(0)$, $T_2^+(0)$, $T_3^+(0)$, $T_1^-(0)$, $T_2^-(0)$, $T_3^-(0)$	0.50,0.50,0.50, 0.50,0.50,0.50

Simulation results for the dynamics of the relative trust for the cultures mentioned in Table 6 are shown in Figure 7.

`



a)



b)

c)

d)

Fig. 7. Dynamics of Relative Trust in Different Cultures. a) Culture A, b) Culture B, c) Culture C, d) Culture D

From Figure 7 it can be concluded that in every culture whatever relative percentage of the positive experiences may be (except when all information agent give negative experiences all of the time (see Figure 7d), the information agent that gives more positive experiences to the human gains more trust. Furthermore, the information agent that gives more positive experiences at least secure neutral trust (*T(t)=0*) in the long run, even the percentage of positive experiences is very low (see Figure 7c).

5 Formal Analysis of The Model

In this section a mathematical analysis is made of the change in trust upon positive (resp. negative) experiences. In Section 2 the differential equation form of the model for positive trust was formulated as:

$$\frac{dT_i^+(t)}{dt} = \beta * \left[\eta * \left(1 - T_i^+(t)\right) + (1-\eta) * \left(\tau_i^+(t) - 1\right) * T_i^+(t) * \left(1 - T_i^+(t)\right)\right] * E_i(t) * \left(E_i(t)+1\right)/2$$
$$- \gamma * T_i^+(t) * \left(1 + E_i(t)\right) * \left(1 - E_i(t)\right)$$

where $\tau_i^+(t)$ is

$$\tau_i^+(t) = \frac{T_i^+(t)}{\sum_{j=1}^n T_j^+(t) / n}$$

One question that can be addressed is when for a given time point *t* an equilibrium occurs, i.e. under which conditions trust does not change at time point *t*. Another question is under which circumstances trust will increase at *t*, and under which it will decrease. As the experience function $E_i(t)$ is given by an external scenario, these questions have to be answered for a given value of this function. So, three cases are considered:

Case 1: $E_i(t) = 1$

In this case the differential equation can be simplified to

$$\frac{dT_i^+(t)}{dt} = \beta * \left(\eta * \left(1 - T_i^+(t)\right) - (1-\eta) * \left(1 - \tau_i^+(t)\right) * T_i^+(t) * \left(1 - T_i^+(t)\right)\right)$$

$$\frac{dT_i^+(t)}{dt} = \beta * \left(\eta - (1-\eta) * \left(1 - \frac{T_i^+(t)}{\sum_{j=1}^n T_j^+(t) / n}\right) * T_i^+(t)\right) * \left(1 - T_i^+(t)\right)$$

It follows that $\frac{dT_i^+(t)}{dt} \geq 0$ if and only if

$$\eta - (1-\eta) * \left(1 - \frac{T_i^+(t)}{\sum_{j=1}^n T_j^+(t) / n}\right) * T_i^+(t) \geq 0$$

or

$$T_i^+(t) = 1$$

For $T_i^+(t) < 1$ this is equivalent to (with: $S(t) = \sum_{j=1}^n T_j^+(t)$)

$$(1-\eta) * \left(1 - \frac{T_i^+(t)}{S(t)/n}\right) * T_i^+(t) \leq \eta$$

$$(1-\eta) * \left(S(t) - n * T_i^+(t)\right) * T_i^+(t) \leq \eta * S(t)$$

$$\left(S(t) * T_i^+(t) - n * T_i^+(t)^2\right) \leq \eta * S(t) / (1-\eta)$$

$$n * T_i^+(t)^2 - S(t) * T_i^+(t) + \eta * S(t) / (1-\eta) \geq 0$$

This quadratic expression in $T_i^+(t)$ has no zeros when the discriminant $S(t)^2 - \frac{4n * S(t) * \eta}{(1-\eta)}$ is negative:

$$S(t)^2 - \frac{4n * S(t) * \eta}{(1-\eta)} < 0 \Leftrightarrow S(t)\left(S(t) - \frac{4n * \eta}{(1-\eta)}\right) < 0 \Leftrightarrow 0 < S(t)/n < \frac{4\eta}{1-\eta}$$

When $\eta > 0.2$ then $1/\eta < 5$ and therefore $1/\eta - 1 < 4$, hence $(1-\eta)/\eta < 4$ which can be reformulated as $4\eta/(1-\eta) > 1$. As $S(t)/n \leq 1$, this shows that for $\eta > 0.2$ as long as $S(t)$ is positive, the discriminant is always negative, and therefore upon a positive experience there will always be an increase. When $S(t) = 0$, which means all trust values are 0, no change occurs. For the case the discriminant is ≥0, i.e., $S(t)/n \geq 4\eta/(1-\eta)$ then the quadratic equation $T_i^+(t)$ for has two zeros symmetric in $S(t)$:

$$T_i^+(t) = \left(S(t) +/- \sqrt{S(t)^2 - \frac{4n * S(t) * \eta}{1-\eta}}\right) \Big/ 2n$$

In this case increase upon a positive experience will take place for $T_i^+(t)$ less than the smaller zero or higher than the larger zero, and not between the zeros. An equilibrium occurs upon a positive experience when $T_i^+(t) = 1$ or when equality holds:

$$n * T_i^+(t)^2 - S(t) * T_i^+(t) - \eta * S(t) / (1 - \eta) = 0$$

This only can happen when the discriminant is not negative, in which case equilibria occur for $T_i^+(t)$ equal to one of the zeros.

Case 2: $E_i(t) = 0$
In this case the differential equation can be simplified to

$$\frac{dT_i^+(t)}{dt} = -\gamma * T_i^+(t)$$

So, in this case positive trust is decreasing or has in equilibrium with positive trust 0.

Case 3: $E_i(t) = -1$
In this case the differential equation can be simplified to

$$\frac{dT_i^+(t)}{dt} = 0$$

So, for this case always an equilibrium occurs in t for positive trust.

For negative trust, the situation is a mirror image of the case for positive trust, and by combining the positive and negative trust, the patterns for overall trust can be analysed.

6 Discussion

This paper has introduced a model for relative trust to enable personal assistant agents to give the appropriate support to humans. Within the model several parameters have been introduced to tailor it towards a particular human. The influence of these parameters upon the trust has been extensively shown in this paper by means of simulations, even including different cultural settings. Finally, a mathematical analysis has been conducted to formally derive what the change of the trust functions is in case of positive and negative experiences.

A variety of trust models have been proposed in the literature [4] and [5]. These trust models attempt to determine the level of trust in certain agents based upon experiences. They do however not take into account the notion of relativeness of this trust. Models have been proposed for relative trust as well. In [9] a model is presented that allows an agent to combine multiple sources for deriving a trust value. This notion of relativeness differs from the notion used in this paper. [10] extends an existing trust model of [11] with the notion of relative trust. They take as a basis certain trust values determined by the model [11], and compare these values in order to make statements about different trust values for different agents. In determining the trust itself, they do not incorporate the experiences with other agents that can perform similar tasks, which is done in this paper. In [12] a trust model is utilized to allocate decision support tasks. In the model, relative trust is addressed as well but again not incorporated in the calculation of the trust value itself.

For future work, an interesting option is to see how well the parameters of the model can be derived by a personal assistant (based upon the requests outputted by the human).

References

1. Aarts, E., Harwig, R., and Schuurmans, M. (2001). Ambient Intelligence. In: P. Denning (ed.), *The Invisible Future*, McGraw Hill, New York, pp. 235-250.
2. Aarts, E.; Collier, R.; van Loenen, E.; Ruyter, B. de (eds.) (2003). *Ambient Intelligence. Proc. of the First European Symposium, EUSAI 2003*. Lecture Notes in Computer Science, vol. 2875. Springer Verlag, 2003, pp. 432.
3. Riva, G., F. Vatalaro, F. Davide, M. Alcañiz (eds.), *Ambient Intelligence*. IOS Press, 2005.
4. Falcone, R., Castelfranchi, C.: Trust dynamics: How trust is influenced by direct experiences and by trust itself. In: Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004), 740–747, 2004.
5. Marx, M., and Treur, J., Trust Dynamics Formalised in Temporal Logic. In: L. Chen, Y. Zhuo(eds.), *Proc. of the Third International Conference on Cognitive Science, ICCS 2001*. USTC Press, Beijing, pp. 359-363, 2001.
6. Edward H. Shortliffe, Bruce G. Buchanan, A model of inexact reasoning in medicine. Mathematical Biosciences, Volume 23, Issues 3-4, April 1975, pp. 351-379
7. Luger, G.F., Stubblefield, W.A., (1998). Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Fourth Edition. Addison-Wesley, 1998, pp. 320-321.
8. Jonker, C.M., and Treur, J., (1999). Formal Analysis of Models for the Dynamics of Trust based on Experiences. In: F.J. Garijo, M. Boman (eds.), *Multi-Agent System Engineering, Proceedings of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'99.* Lecture Notes in AI, vol. 1647, Springer Verlag, Berlin, 1999, pp. 221-232.
9. Beth, T., Borcherding, M., and Klein, B., “Valuation of trust in open networks”, Proceeding of 3rd European Symposium on Research in Computer Security (ESORICS), pp. 3-18, 1994.
10. Kluwer, J. and Waaler, A., Relative Trustworthiness, In: Dimitrakos, T. *et al.* (eds), FAST 2005, pp.158-170, 2006.
11. Jones, A., On the concept of trust, *Decision Support Systems*, 33:225-232, 2002.
12. Maanen, P.-P. van, Dongen, K. van, Towards Task Allocation Decision Support by means of Cognitive Modeling of Trust, In: C. Castelfranchi, S. Barber, J. Sabater, and M. Singh (Eds.), *Proceedings of the Eighth International Workshop on Trust in Agent Societies (Trust 2005)*, pp. 168-77, 2005.