

# Trust Dynamics Formalized in Temporal Logic

Maarten Marx and Jan Treur  
*Department of Artificial Intelligence, Vrije Universiteit Amsterdam,  
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands  
Email: treur@cs.vu.nl URL: <http://www.cs.vu.nl/~treur>*

**In this paper the dynamics of trust in the light of experiences is analysed, and formalized in temporal logic. This logic can be used for the analysis and specification of models for trust evolution and trust update. It is shown for a number of properties of models for trust dynamics how they can be formalized; also some relevant properties are shown that cannot be formalised in temporal logic.**

## 1. INTRODUCTION

Trust is the attitude an agent has with respect to the dependability/capabilities of some other agent (maybe itself) or with respect to the turn of events. The agent might for example trust that the statements made by another agent are true. The agent might trust that another agent (or itself) is capable of performing certain tasks. In (Castelfranchi and Falcone, 1998a, 1998b) the importance of the notion trust is shown for agents, multi-agent systems, and their foundations. (Elofson, 1998) states that the reach and effect of trust in the affairs of individuals and organizations is largely pervasive. The definition of (Lewis and Weigert, 1985) refers to observations which in turn lead to expectations: ‘observations that indicate that members of a system according to and are secure in the expected futures constituted by the presence of each other for their symbolic representations’. (Elofson, 1998) agrees that observations are important for trust, and he defines: ‘trust is the outcome of observations leading to the belief that the actions of another may be relied upon, without explicit guarantee, to achieve a goal in a risky situation’.

In (Jonker and Treur, 1999) the dynamics of trust in the light of experiences is analyzed formally. In this paper no commitment or limitation was made to a specific logical language; the properties were expressed in general mathematically defined terms such as trust evolution functions and trust update functions. For the purpose of expressivity, the question whether or how the properties can be formalized in an appropriate logical language was ignored. Therefore, different possible properties of these models for trust dynamics have been identified and defined mathematically for which it is not clear a priori whether they can be formalised in available logical languages as well. Given these properties that have been identified and mathematically defined, in the current paper the question of whether or not (and how) they can be expressed in a logical language is considered.

After the preliminaries in Section 2, in sections 3 and 4 a number of properties of trust evolution functions and trust update functions as introduced in (Jonker and Treur, 1999) are inspected. It is shown how some of them can be formalized in temporal logic, and why some cannot be formalized in temporal logic. Section 5 concludes the paper.

## 2. PRELIMINARIES

In this paper, trust is considered a mental agent concept that depends on experiences. This is modeled by a function that relates sequences of experiences to trust representations: a *trust evolution* function. Another modeling is given by a recursively defined function relating a current trust representation and a current experience to the next trust representation: a *trust update* function. To obtain a formal framework, the following four sets are introduced. A partially ordered set  $\mathbb{E}$  of *experience values*, the set  $\mathbb{N}$  of natural numbers, the set  $\mathbb{ES}$  of *experience sequences*, i.e., functions from  $\mathbb{N}$  to  $\mathbb{E}$ , and a partially ordered set  $\mathbb{T}$  of *trust values*. Within the sets  $\mathbb{E}$  and  $\mathbb{T}$  subsets of positive and negative values are distinguished. Examples of such sets are the (closed) interval of real numbers between -1 and 1, or the set  $\{-1, 0, 1\}$ . A *trust evolution function* is a function  $te : \mathbb{ES} \times \mathbb{N} \rightarrow \mathbb{T}$ . A *trust update function* is a function  $tu : \mathbb{E} \times \mathbb{T} \rightarrow \mathbb{T}$ . Throughout this paper we assume that the value of a trust evolution function  $te$  at time  $t$  only depends on the experiences in the past of  $t$  (*future independency*).

For a temporal logic formalisation, *state atoms* are atoms (such as  $ev(v)$  and  $tv(w)$ ) that express the experience value  $v$  and trust value  $w$  in a state, and relations for these values such as  $<$  or  $pos$ . A *state* is a truth assignment (of truth values  $true$  and  $false$ ) to the set of (ground) state atoms;  $s$  denotes the set of all states. A *temporal model* is a function  $M : \mathbb{N} \rightarrow s$ . Within the temporal logic the following *temporal operators* are used:  $X$  with  $Xa$  meaning that  $a$  holds in the next instant of time;  $F$ , with  $Fa$  meaning that  $a$  will be true at some time point in the future; and  $G$ , with  $Ga$  meaning that  $a$  will be true at all future time points. Repetition ( $n$  times) of an operator such as  $X$  is denoted by  $X^n$ . For more details about temporal logic, see (Burgess, 1984).

## 3. PROPERTIES OF TRUST EVOLUTION FUNCTIONS EXPRESSED IN TEMPORAL LOGIC

In (Jonker and Treur, 1999) a number of interesting possible properties of trust evolution functions are defined. For a sample of these it is investigated here whether they can be formalized in discrete linear time temporal logic. For motivation and further explanation we refer to the reference mentioned. In stating the properties we implicitly universally quantify over experience sequences  $e$ . The following properties from (Jonker and Treur, 1999) are considered (here  $e, f \in \mathbb{ES}$ ,  $s, t, u \in \mathbb{N}$ ).

### *Future independence*

Future independence expresses that trust only depends on past experiences, not on future experiences. This is a quite natural assumption that is assumed to hold for all trust evolution functions. In mathematical terms:

$$\text{if } e|_{s:t} = f|_{s:t} \quad \text{then} \quad te(e, t) = te(f, t)$$

This property refers to two different histories and compares them. Since in linear time temporal logic only a reference can be made to time points in one history, this property cannot be expressed.

### *Monotonicity*

Monotonicity expresses that the more positive the experiences are, the higher the trust. In mathematical terms:

$$e \leq f \Rightarrow te(e, t) \leq te(f, t)$$

Also this property refers to two different histories and compares them, and therefore cannot be expressed in temporal logic.

### *Positive trust extension*

Positive trust extension expresses that trust is increasing if only positive experiences are encountered, i.e., after a positive experience, trust will become at least as much (less) as it was. In mathematical terms:

$$\forall s, t [\forall u \in N : s \leq u < t : e_u \text{ positive}] \Rightarrow te(e, s) \leq te(e, t)$$

This property can be expressed in temporal logic by the following scheme (that can be instantiated for all experience values  $v$  and trust values  $w$  and  $w'$ ):

$$ev(v) \wedge pos(v) \wedge tv(w) \wedge Xtv(w') \rightarrow w \leq w'$$

Explanation: if at a point in time the experience value is  $v$  (expressed by the atom  $ev(v)$ ) and  $v$  is positive (expressed by  $pos(v)$ ), and the trust value is  $w$  (expressed by  $tv(w)$ ), and the next trust value is  $w'$  (expressed by the temporal next-operator  $X$  followed by the atom  $tv(w')$ ), then  $w'$  is at least as high as  $w$ .

### *Degree of trust dropping $n$*

The property degree of trust dropping (or gaining) expresses after how many negative (or positive) experiences trust will be negative (or positive). In mathematical terms:

$$\forall t [\forall k \in N : t-n < k \leq t : e_k \text{ negative}] \Rightarrow te(e, t) \text{ negative}$$

This property can be expressed in temporal logic by the following scheme:

$$ev(v_0) \wedge neg(v_0) \wedge X(ev(v_1) \wedge neg(v_1)) \wedge \dots \wedge X^{n-1}(ev(v_{n-1}) \wedge neg(v_{n-1})) \wedge X^{n-1}(tv(w)) \rightarrow neg(w)$$

Explanation: if at a point in time the experience value is some  $v_0$  and  $v_0$  is negative, and the same holds for the next  $n-1$  time points, and the trust value after  $n-1$  time points is  $w$ , then this  $w$  is negative.

### *Positive limit approximation*

Positive limit approximation expresses that it is always possible to reach maximal trust, if a sufficiently long period with only positive experiences is encountered (and the same for the negative case). In mathematical terms:

If there exists an  $M$  such that for all  $s > M$  it holds that  $e_s$  is maximal (in  $E$ ), then an  $N$  exists such that  $te(e, t)$  is maximal (in  $T$ ) for all  $t > N$ .

This property can be expressed in temporal logic by the following scheme:

$$FG(ev(v) \rightarrow \bigwedge_{v'} \neg v' > v) \rightarrow FG(tv(w) \rightarrow \bigwedge_{w'} \neg w' > w)$$

Here  $\bigwedge_{v'}$  stands for the conjunction over all experience values  $v'$  and  $\bigwedge_{w'}$  for the conjunction over all trust values  $w'$ . Explanation: if a future time point exists such that for all later time points if  $v$  is the experience value, then no (experience) value is higher than  $v$ , then a future time point exists such that for all later time points if  $w$  is the trust value, then no (trust) value is higher than  $w$ .

#### 4. PROPERTIES OF TRUST UPDATE FUNCTIONS IN TEMPORAL LOGIC

In (Jonker and Treur, 1999) also a number of possible properties of trust update functions are defined and related to properties of trust evolution functions. Here we list some of them, formalize some of them in temporal logic and show how it can be used to reason formally about trust update functions.

##### *Monotonicity*

We call a trust update function  $tu$  monotonic if higher experience values and higher trust values lead to higher trust update values. In mathematical terms:

$$ev1 \leq ev2 \ \& \ tv1 \leq tv2 \quad \Rightarrow \quad tu(ev1, tv1) \leq tu(ev2, tv2)$$

Just as monotonicity for evolution functions this property is not expressible in temporal logic, since comparison of two states in different histories is involved.

##### *Positive and negative trust extension*

This property states that positive (negative) experiences lead to higher (lower) trust values. In mathematical terms:

$$ev \text{ positive} \Rightarrow tu(ev, tv) \geq tv$$

This property can be expressed in temporal logic by the following scheme:

$$ev(v) \wedge pos(v) \wedge tv(w) \wedge Xtv(w') \rightarrow w' \leq w$$

Notice that this is the same formalisation as the one for positive trust extension for trust evolution functions.

##### *Strict positive (negative) monotonic progression*

This property is a stronger version of the previous one and states that positive (negative) experiences lead to strictly higher (lower) trust values, as long as this is possible. In mathematical terms:

$$ev \text{ positive and } tv \text{ not maximal (in } T) \Rightarrow tu(ev, tv) > tv$$

This property can be expressed in temporal logic by the following scheme:

$$ev(v) \wedge pos(v) \wedge tv(w) \wedge Xtv(w') \rightarrow (w' > w \vee \bigwedge_{w''} \neg w'' > w)$$

Explanation: if at a point in time the experience value is  $v$  and  $v$  is positive, and the trust value is  $w$ , then the next trust value  $w'$  is higher than  $w$  or no trust values higher than  $w$  exist.

The following proposition exemplifies how properties of trust update functions and trust evolution functions can be related. Based on the formalisations introduced above, it can be proven within temporal logic.

#### **Proposition**

Let  $tu$  be a trust update function,  $it$  some initial trust value and  $te$  the trust evolution function generated by  $tu$  and  $it$ , i.e.,

$$\begin{aligned} te(e,0) &= it && \text{for all } e \in ES \\ te(e, t+1) &= tu(e, te(e, t)) && \text{for all } e \in ES, t \in \mathbb{1} \end{aligned}$$

If the set of trust values  $\tau$  is finite and  $tu$  satisfies strict positive monotonic progression, then  $te$  has positive limit approximation.

## 5. CONCLUSIONS AND FURTHER RESEARCH

To formalize and analyse dynamic phenomena, often it is implicitly or explicitly claimed that temporal logic, e.g., linear time temporal logic, is useful; e.g., (Burgess, 1984; Fisher, 1994; Fisher and Wooldridge, 1997). To evaluate such a claim, this paper investigates the expressibility of a number of possible dynamic properties of trust dynamics, as identified earlier and formalised in general mathematical terms, in (discrete linear time) temporal logic. The result of this investigation is partly positive and partly negative. Indeed, it turns out that a number of relevant properties of trust dynamics can be expressed. However, also a number of relevant properties cannot be expressed in standard temporal logic. Although primarily the use of linear time temporal logic was investigated, the reason for this lack of expressivity (i.e., the impossibility to refer to and compare different histories) turns out to be rather fundamental; for example, for the same reason, also branching time temporal logic will fail to express these properties. In further research it will be investigated whether and how dynamic properties of trust can be expressed in a logical language in which explicit reference can be made to different traces of histories.

## REFERENCES

- Burgess, J.P. (1984). Temporal logic. In: D.M. Gabbay and F. Guenther, (eds.), *Handbook of Philosophical Logic*, vol. 2. Reidel, Dordrecht.
- Castelfranchi, C., and Falcone, R. (1998a). Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification. In: Demazeau, Y. (ed.), *Proc. of the Third International Conference on Multi-Agent Systems*, IEEE Computer Society, pp. 72-79.
- Castelfranchi, C., and Falcone, R., (1998b). Social Trust: Cognitive Anatomy, Social Importance, Quantification, and Dynamics. In: *Proc. of the First International Workshop on Trust*, pp. 35-49.
- Elofson, G., Developing Trust with Intelligent Agents: An Exploratory Study. In: *Proc. of the First International Workshop on Trust*, 1998, pp. 125-139.
- Fisher, M. (1994). A survey of Concurrent MetateM — the language and its applications. In: D.M. Gabbay, H.J. Ohlbach (eds.), *Temporal Logic — Proc. of the First International Conference*, Lecture Notes in AI, vol. 827, pp. 480–505.
- Fisher, M., and M. Wooldridge (1997). On the formal specification and verification of multi-agent systems. *International Journal of Co-operative Information Systems*, IJCIS vol. 6 (1), special issue on Formal Methods in Co-operative Information Systems: Multi-Agent Systems, Huhns, M. and Singh, M. (eds.), pp. 37–65.
- Jonker, C.M., and Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In F.J. Garijo and M. Boman, (eds.), *Multi-Agent System Engineering*, Proc. of MAAMAW'99, Lecture Notes in AI, vol. 1647, Springer Verlag, pp. 221-232.
- Lewis, D., and Weigert, A. (1985). Social Atomism, Holism, and Trust. In: *Sociological Quarterly*, pp. 455-471.