The Impact of Benevolence in Computational Trust

Joana Urbano, Ana Paula Rocha, and Eugénio Oliveira

LIACC / DEI, Faculdade de Engenharia, Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
{joana.urbano,arocha,eco}@fe.up.pt

Abstract. Trust is a construct of paramount importance in society. Accordingly, computational trust is evolving fast in order to allow trust in artificial societies. Despite the advances in this research field, most computational trust approaches evaluate trust by estimating the trustworthiness of the agents under evaluation (the trustees), without however distinguishing between the different dimensions of trustworthiness, such as ability and benevolence. In this paper, we propose different techniques to extract the ability of the trustee in the task at hand and to infer the benevolence of the trustee toward the truster when the trust judgment is made. Moreover, we propose to dynamically change the relative importance and impact of both ability and benevolence on the perceived trustworthiness of the trustee, taking into consideration the development of the relationship between the truster and the trustee and the disposition of the truster in the specific situation. Finally, we set an experimental scenario to evaluate our approach. The results obtained from these experiments show that the proposed techniques significantly improve the reliability of the estimation of the trustworthiness of agents.

Keywords: computational trust, benevolence, trustworthiness.

1 Introduction

Computational trust is considered an enabler technology in virtual societies, and the estimation of trustworthiness is paramount to assess the trust that a truster agent has on a given trustee. An individual is more or less trustworthy in performing a task in a given situation depending on his ability in the matter, his overall integrity, and the stage of his relationships with the truster. Therefore, in order to better estimate the trustworthiness of agents, it is important to consider these three dimensions individually, and to combine them in a dynamic way taking into consideration the situation and the development of the relationship. However, the majority of the computational trust approaches presented in literature estimates the trustworthiness of agents as a block and does not distinguish between these trustees' attributions.

In this paper, we present a computational trust approach grounded on multidisciplinary literature on trust that is able to capture the ability and benevolence

C.I. Chesñevar et al. (Eds.): AT 2013, LNAI 8068, pp. 210–224, 2013.

of the agent under evaluation. Through its main component, *Social Tuner*, our approach novels in its ability to estimate the trustee's benevolence at the moment of the trust decision from the evidence available on this trustee.[1] Moreover, our approach combines the estimated ability of the trustee with his estimated benevolence, as computed by *Social Tuner*, into a trustworthiness score, where the relative importance and impact of ability and benevolence take into consideration the development of the relationship between truster and trustee at the time of the assessment. To prove the benefits of our benevolence-based computational trust approach, we enhanced three known trust-based evidence aggregators – the one described in the Beta Reputation model [2], the asymmetry-based trust update function described in [3], and our model Sinalpha [4] –, by adding the functionalities of *Social Tuner* to these aggregators. The results we obtained and present in this paper are very encouraging, as they showed that there is a clear benefit in using *Social Tuner* in the described situations: the benevolence-enhanced trust models allowed for a more accurate estimation of the trustees' trustworthiness than the original computational trust models.

This paper is organized as follows: in the next section, we overview theoretical concepts relating trust and trustworthiness with benevolence. In Section 3, we present the related work. Section 4 presents the main motivation for our work and basic notation. In Section 5, we present our computational trust approach, which is experimentally evaluated in Section 6. Finally, Section 7 concludes the paper and presents future work.

2 The Relation between Trust and Benevolence

Trust is a property of the one that trusts, the truster, in relation to the object of this trust, the trustee. To infer the trust on others, the truster needs to estimate the trustworthiness of these others ([5–8]). In the same way, the truster's propensity to trust ([9, 6–8]), his emotional state ([10]), the trustee's physical and cultural characteristics ([11]), and potentially his reputation ([2, 11, 8]), are other factors that the truster normally weights when making a trust judgment. However, in this paper, we focus on the role of trustworthiness on trust.

A trustworthy entity would normally present high values of ability, integrity and benevolence in the situation under assessment ([9, 12, 7]), and his behavior would be predictable in this situation ([11, 8]). Ability relates to the potential competence of the evaluated entity to do a given task, and is probably the trustworthiness dimension most mentioned by trust scholars (e.g., [9, 5, 13, 14, 8, 15]). The truster perceives the trustee's qualities that make the trustee able for the task (e.g., skills, know how, general wisdom, self-esteem, self-confidence, and leadership) as mainly a cognitive process and less of an emotion-based process ([7]). Integrity and benevolence, however, are often overlooked by scholars, particularly computer scientists addressing the trust topic. In this paper, we are particularly interested on benevolence, and do not further address the integrity construct. Next, we overview essential theoretical aspects of benevolence.

[1] An early draft of our work on benevolence-based trust is presented in [1].

2.1 Benevolence

Benevolence is considered by several scholars as a key element of close relationships and an antecedent of trustworthiness (e.g., [9, 12, 13, 16–18]). Benevolence is either a disposition to do good and an act of kindness, where the trustee has a feeling of goodwill toward the interacting partner excluding any intention of harming him given the opportunity to do so ([12, 13]). It usually implies a specific attachment of the truster toward the trusted one, excluding any motivation based on egocentric profit motives (e.g., [9, 12]). Different studies on individual differences and human behavioral genetics link benevolence to Agreeableness and Neuroticism ([19, 20]), two personality traits that are influenced by heredity, environment, time and gender ([21, 22]). Recent advances in the area of behavioral neurology and cognitive neuroscience relate the human amygdala with expressions of benevolence and normal interpersonal trust ([18]). Benevolence is also being positively correlated with the recognition of kinship and physical resemblance (e.g., [23, 17]) and with in-group awareness and cultural relatedness (e.g., [13, 16, 24]). All these studies propose that individuals have a disposition toward benevolence, with some individuals being more benevolent than others in identical situations.

Proposition 1. DISPOSITION TO BENEVOLENCE*: Each individual has a specific disposition to benevolence, related with his traits of personality.*

Benevolence also develops in long-term and close relationships, where trust is reciprocated and positive affect circulates among those who express trust behaviorally, which may result in intense emotional investments being made ([25]) and in the internalization of relational norms and values ([16]).

Proposition 2. RELATIONAL BENEVOLENCE*: In long-term and close relationships, affective commitment arises and has a positive impact on the benevolence of partners. Then, the benevolence of the partner is usually perceived much later in the relationship than this partner's ability.*

Some authors consider that there is a different form of benevolence ('mutualistic' benevolence) that is motivated by the expectation of joint gain ([23, 26, 16]), where the voluntary helping behaviors beyond the call of duty still exists. Most partners that establish ongoing trust relationships benefit from the benevolent actions of the other partner, and tend to act benevolently in order to maintain the relationship and continue profiting from these trust-based benefits (e.g., [12, 27, 26, 28, 16]). In these relationships, the satisfaction of partners increase with the perception of the equity in the exchange and the perception of continuity of the relationship ([12, 28, 16]). The partners probably do not risk investing in the development of new relationships if they already have several ongoing relationships ([26]). In the same way, the value they attach to a given trust relationship may diminish if they perceive that the likelihood of being trusted by somebody else is high ([12]). If we add to the satisfaction with the relationship some form of utilitarianism – where individuals are more willing to rely on partners when they expect that the interaction with these partners brings more benefits than costs [28] – then we consider that the partners to the exchange developed a calculative commitment that eventually leads to the mutualistic form of benevolence ([16]).

3 Related Work

The work in [15] presented a conceptual model of social trust based on [11] that distinguishes between ability, positive intentions, ethics, and predictability. The authors suggested a probabilistic approach to implement the model but recognized the limits of such an approach in the treatment of the cognitive and social concepts involved; this model was not implemented.

To date, the only computational approach that included a comprehensive set of features grounded on the theory of trust and that was actually implemented is the socio-cognitive model of trust by Castelfranchi and Falcone (e.g., [8, 29]). This model considers that the truster has a goal that can be accomplished by an action of the trustee, and that trust in a particular situation is formed by considering the different beliefs that the truster has about the trustee, either internal (beliefs on competence, disposition, and unharmfulness) or external (opportunities and dangers). The values of these beliefs are further modulated by meta-beliefs about the relative strength of each belief. The richness of this model makes it hard to implement in practice. In fact, the current implementation of the model (e.g., [8, 29]) requires extensive manual configuration by domain experts for each trustee and task under assessment and oversimplifies the theoretical model. Moreover, it requires explicit information about the competence and disposition (or similar beliefs) of the agent under evaluation, which may be hard to get in dynamic agent-based environments. In our model, we adopted a more pragmatic approach in the sense that we consider that the available evidence may be scarce and does not necessarily discriminate about the different attributions of the trustee. Finally, the work in [30] formalized in multimodal logic the model of Castefranchi and Falcone, adding the notions of occurrent trust and dispositional trust (i.e., trust in a general disposition of the trustee to perform a similar task some point in the future).

4 Motivation and Notation

Most computational trust approaches estimate the trustees' trustworthiness using individual items of evidence about these trustees' behavior in past interactions, either with the truster or with third party agents. However, none of these approaches is able to estimate the benevolence of the trustee toward the truster from the set of past evidence. Nevertheless, the particular outcome of an exchange may depend not only on the ability (and integrity, predictability, etc) of the trustee, but also on the benevolence relationship that exists between the latter and the truster. In fact, we believe that understanding the benevolence of the trustee toward the truster at the moment of the trust decision is fundamental to accurately estimate the latter's trustworthiness. With this in mind, we present the main hypothesis of this work, as follows.

Hypothesis 1. *The extraction of benevolence-based information from the set of evidence on the trustee under evaluation and its use in adequate stages of the relationship between truster and trustee improves the reliability of the estimation of this trustee's trustworthiness. The consequent reliability of the trust decision is improved even when the available evidence is scarce.*

4.1 Basic Notation

Our generic computational trust model is applied to environments where truster agents select the best trustees to interact with, with the posterior establishment of dyadic agreements between partners. We assume the existence of a society of agents represented by the limited set $\mathcal{A} = \{a_1, a_2, ..., a_n\}$. In this society, agents make trust decisions about other agents concerning the realization of a given task $t \in \mathcal{T}$ in a given situation $s \in \mathcal{S}$, where $\mathcal{T} = \{t_1, t_2, ..., t_m\}$ is the set of all possible m tasks in the society and $\mathcal{S} = \{s_1, s_2, ..., s_k\}$ is the set of all possible k situations in the society. In order to characterize and describe the situation leading to an agreement, we consider the definition of context as including four main types of context: identity, time, location, and activity [31]. Furthermore, we consider that context is expressed by eight dimensions $d_1, d_2, ..., d_8$, where dimensions d_1 and d_2 identify the truster and the trustee of the reported interaction, respectively; d_3 and d_4 represent the time and location of agreement; and d_5, d_6, d_7, and d_8 identify and characterize the type of the task, its complexity, deadline and outcome of its realization, respectively.

In this work, we assume that all agreements performed in the society of agents refer to the same type of task t (d_5), although it can assume different degrees of complexity (d_6) and deadlines (d_7). Also, we consider that the set of possible outcomes (d_8) is defined by $\mathcal{O} = \{F, Fd, V\}$, where F (fulfill) means that the truster considers that the trustee performed whatever matter he had to perform on time, Fd (fulfill with delay) means that the truster was presented with an unexpected delay in the realization of the task, and V (violation) means that the truster considers that the trustee presented a severe contingency (e.g., the task was not even performed, or the delay was excessive, or the quality was way below the acceptable). We further consider that the relative preference relations over these values is given by $F \succ Fd \succ V$ (i.e., F is strictly preferable over Fd, and Fd is strictly preferable over V), for all agents of the society.

In the sequence of our characterization of context, we represent any situation $s_i \in \mathcal{S}$ as a tuple of values ascribed to each contextual dimension but the one corresponding to the outcome dimension: $s_i = \langle v_1^{s_i}, v_2^{s_i}, ..., v_7^{s_i} \rangle$., where $v_j^{s_i}$ is the value ascribed to dimension j in situation s_i. Similarly, an individual item of evidence e_i is also represented using a tuple of values ascribed to each contextual dimension, but now the outcome o^{e_i}, corresponding to dimension d_8, is mandatory: $e_i = \langle v_1^{e_i}, v_2^{e_i}, ..., v_8^{e_i} \rangle$. Finally, the set of all items of evidence existing about a given trustee y is given by $E_{*,y} = \{e_i \in \mathcal{E} : v_2^{e_i} = y\}$, where $\mathcal{E}$ represents all evidence available on all agents of society $\mathcal{A}$. In the same way, $E_{x,y}$ represents all evidence about the direct past experiences of truster x with trustee y, such that $E_{x,y} = \{e_i \in \mathcal{E} : v_1^{e_i} = x, v_2^{e_i} = y\}$.

5 Our Computational Model of Trust

The benevolence-based computational model of trust that we present in this paper is part of a larger framework of social trust that we are developing at our Laboratory. It integrates three distinct functions: the ability evaluation function ($A_{x,y} : \ \mathcal{S} \times E_{*,y} \rightarrow [0,1]$), the benevolence evaluation function ($B_{x,y} : \ E_{x,y} \rightarrow [0,1]$), and the trustworthiness evaluation function ($Tw_{x,y} : \ [0,1] \times [0,1] \rightarrow [0,1]$). We describe each of these functions (whose relation is illustrated in Figure 1) in the following subsections.

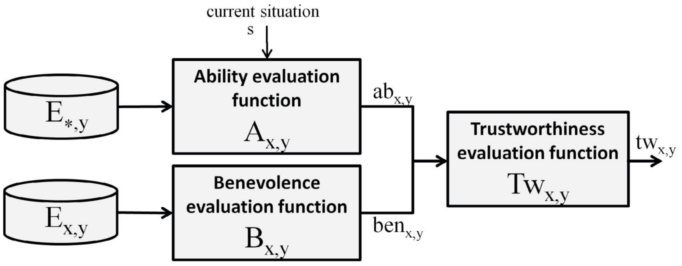

Fig. 1. Our benevolence-based computational model of trust

5.1 Function $A_{x,y}$ – The Ability Component

As mentioned in the previous section, the ability evaluation function takes as input all evidence available on the trustee ($E_{*,y}$) and the representation of the situation s under assessments (cf. Figure 1). Taking into consideration that our representation of evidence makes no explicit reference to the ability of the trustee, we infer his ability from the aggregation of all evidence available, hoping at least to understand whether he has very low ability (tending to violate most of his agreements) or very high ability (tending to fulfill most of his agreements). To this purpose, several existing trust-related evidence aggregators may be used, such as the ones described in [2, 32, 3, 4].

5.2 Function $B_{x,y}$ – The Social Tuner Component

The *Social Tuner* component is our proposal to instantiate the benevolence evaluation function $B_{x,y}$ represented at the beginning of this section. Similarly to what happened when defining function $A_{x,y}$, the challenge that we face is to extract any information available about the trustee's benevolence toward the truster using only the structured, simple data from the evidential set $E_{x,y}$. In this respect, we hypothesized that the use of such information would improve the reliability of the trustworthiness estimation (Hypothesis 1). Of course, we realize that any approach to benevolence in such conditions could not be comprehensive in covering the benevolence concept. However, we believe that our initial purpose of getting more from the available set of evidence in order to increase the reliability of the estimated trustworthiness still maintains its validity.

In particular, *Social Tuner* measures the trustee's specific attachment toward the truster, i.e., his disposition to do good to the truster. This is captured by the *coefficient of benevolent actions* parameter, which we present next.

Coefficient of Benevolent Actions. The *coefficient of benevolent actions*, $\rho_{ba} \in [0,1]$, measures the trend of contingencies presented by the trustee to the truster in the past. In this paper, the truster considers that the outcome Fd corresponds to a mild contingency, while V is perceived as a severe contingency. Hence, the first step to calculate the trend of contingencies is to define how much the truster values each outcome of his possible agreements, using function $vl : \mathcal{O} \rightarrow [0,1]$. Here, we consider that $vl(F) = 1.0$, $vl(Fd) = 0.5$, and $vl(V) = 0.0$.

Then, we build a function of the cumulative value of past agreements per generated outcome, *cumValAgreem*, which we define in Equation 1. Figure 2 (*left*) illustrates the cumulative values of outcomes' curve of three different trustees, each one having interacted 10 times with a given truster in the past, where one of them fulfilled all agreements with the truster, the other delayed all the agreements, and the remaining violated all agreements.

$$cumValAgreem(i) = \sum_{j=1}^{i} vl(o^{e_j}) \ . \tag{1}$$

Finally, the *coefficient of benevolent actions* is given by the correlation between the number of agreements established between truster and trustee in the past and the function of the cumulative value of past agreements calculated for these agents. In order to get this correlation, we apply a linear regression to the function of the cumulative value of past agreements. Figure 2 (*right*) illustrates this process for two different agents: one that is initially very observant of his obligations toward the truster but that inverts this behavior in the last agreements, and the other presenting the opposite behavior.

Reminding the linear regression function for one predictor, $Y = B_0 + B_1.X$, we use this function to indicate the progress of the cumulative value of past agreements, where X represents the past agreements and Y the cumulative function. Particularly, we use the intercept (B_0) and the regression coefficient (B_1) to estimate if the trustee's benevolence toward the truster is steady, is progressing positively, is progressing negatively, etc. This means that by using this process we are able to estimate how the benevolence of this relationship is evolving.

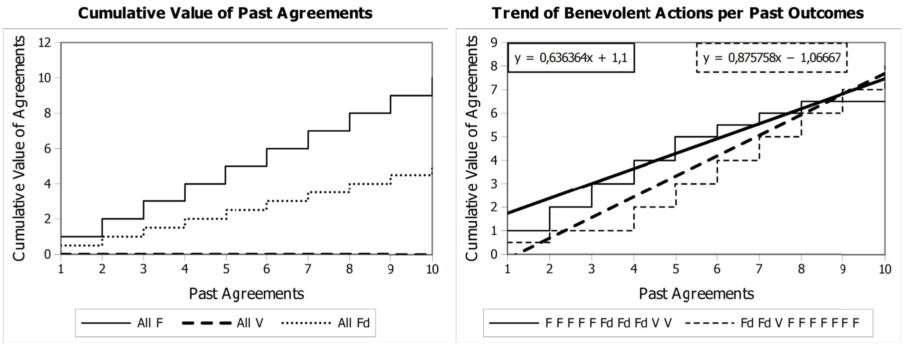

Fig. 2. Instances of the functions of cumulative value of agreements (*left*) and of benevolent actions per past agreements (*right*)

Finally, the coefficient of benevolent actions is given by a function of the correlation coefficient and the intercept, as illustrated in Equation 2.

$$\rho_{ba} = B_0 + 0.10 B_1 \ . \tag{2}$$

The value of this coefficient is minimum ($\rho_{ba} = 0$) when the trustee constantly delivered the worse possible outcomes (i.e., V) in past agreements with the truster indicating that he was acting with no benevolence at all toward the truster. Conversely, this value is maximum when the trustee totally fulfilled all the past agreements with the truster, showing high benevolence toward him.

Estimating the Trustee's Benevolence. The estimated value of the benevolence of the trustee toward the truster, $ben_{x,y}$, is derived from the *coefficient of benevolent actions* using the formula in Equation 3.

$$ben_{x,y} = \frac{1}{2}\rho_{ba} + \frac{1}{2}\frac{\sum_{i}^{|E_{x,y}|} vl(o^{e_i})}{|E_{x,y}|} \ . \tag{3}$$

It is worth noting that the estimation of benevolence is only possible when there are, at least, two past interactions between the truster and the trustee under evaluation. In the same way, this estimated value of the benevolence must be updated at every new trustworthiness estimation, as the benevolence of agents may evolve due to the mutualistic satisfaction/dissatisfaction of the trustee with the relationship, which may change with time and context.

By evaluating the benevolence of the trustee toward the truster, we are able to account for the *emotional content of trust*. For example, let us imagine that traditional (single-dimension) evidence aggregator derived a low to medium value of trustworthiness for the trustee under evaluation; this might indicate that the trustee is low in ability, benevolence, or both. However, if the *Social Tuner* indicates a high value of benevolence of the trustee toward the truster, this may mean that both partners are engaged in a benevolent relationship, and that the truster may expect the trustee to fulfill a future joint agreement.

In a contrasting example, if *Social Tuner* detects a low benevolence level toward the trustee and the general trustworthiness score of the latter is high, it is highly probable that the trustee has high ability in performing the task, but has low benevolence toward the truster. Knowing this information, the truster can either avoid to enter in a future agreement with the trustee, or give the first step to promote goodwill trust by not denouncing a contingency by the trustee. However, if the trustee's trustworthiness is low, this might indicate that the trustee is either very low in ability or very low in benevolence (or both cases), which gives a precious clue to the truster that the trustee is possibly not a good partner to establish an agreement with.

5.3 Function $Tw_{x,y}$

The trustworthiness evaluation function $Tw_{x,y}$ takes into consideration the perception of the ability and benevolence of the trustee, ascribing more weight to the ability dimension when both truster and trustee are practically strangers, and progressively increasing the weight of benevolence as the partners get to know each other better (Proposition 2). $Tw_{x,y}$ is shown in Algorithm 1.

Algorithm 1. Function $Tw_{x,y}$

```
 1: function TW ($E_{*,y}, N_{ben_{close}}$) returns $tw_{x,y}$
 2:     $E_{*,y}$: the set of all evidence about trustee $y$
 3:     $N_{ben_{close}}$: minimum $(x, y)$ interactions for closeness
 4:
 5:     $E_{x,y} \leftarrow \{e^i \in E_{*,y} \ : \ v_1^{e_i} \neq x\}$
 6:     $ab_{x,y} \leftarrow$ Ability $(E_{*,y})$
 7:     $ben_{x,y} \leftarrow$ Social Tuner $(E_{x,y})$
 8:     $N_{x,y} \leftarrow |E_{x,y}|$
 9:     if $N_{x,y} > N_{ben_{close}}$ then $N_{x,y} = N_{ben_{close}}$
10:     if $N_{x,y} > 1$ then $\omega_{ben} = N_{x,y}/N_{ben_{close}}$
11:     else $\omega_{ben} = 0$
12:     $tw_{x,y} = (1 - \omega_{ben}) \cdot ab_{x,y} + \omega_{ben} \cdot ben_{x,y}$
13: return $tw_{x,y}$
```

In the algorithm, we measured the number of interactions between x and y, $N_{x,y}$ (line 8), and defined a minimum number of interactions between truster x and trustee y, $N_{ben_{close}}$, after which the partners are considered to be engaged in a close relationship (lines 3 and 9). Also, we considered a weight of benevolence, ω_{ben}, to be used when combining the estimated value of the trustee's ability as returned by *Ability* (line 6) with the estimated value of its benevolence as returned by *Social Tuner* (line 7). This weight is set to zero when there is just one or zero interactions between both partners (line 11), and then progressively increases with the growing number of interactions between the partners, until it reaches the maximum value of one when the partners are considered to be in a close relationship (line 10). Finally, the estimated value of the trustee's trustworthiness ($tw_{x,y}$) is computed using the weighted mean of $ab_{a,y}$ and $ben_{x,y}$ with weights $(1 - \omega_{ben})$ and ω_{ben} (line 12).

6 Simulated Experiments

6.1 Experimental Design

The experiments were conducted in an agent-based simulated environment where, at every round of the simulations, different types of trusters chose the best partners to perform a task from a set of trustees with different characteristics. For simplicity, we considered that there was only one task being negotiated by all trusters, although its requirements in terms of complexity and due time changed with round and truster; also, all trustees accepted to negotiate with all trusters. We used a behavioral model of agents that we have developed in [1]. This model starts after the establishment of an agreement between the truster and the selected trustee, thus excluding the selection process itself. It focuses on both types of agents' decision concerning the fulfillment of the established agreement: the trustees may opt to fulfill the agreement (trusters will report outcome F), or to delay its realization; accordingly, the trusters may respond to a delay by either retaliating, denouncing the breach (reporting outcome V), or forgiving the contingency (reporting outcome Fd).

Decision about the Agreements' Outcome. The final decision about the outcome of the agreement is governed by different parameters, including the disposition to benevolence (Proposition 1) and the ability of agents, the satisfaction with the relationship, and the mutualistic benevolence of agents. On one hand, every agent was assigned a random dispositional benevolence at setup following an uniform distribution over values *low*, *medium* and *high*. Besides, agents with the role of trustee were also randomly assigned a value of ability following a similar distribution. On the other hand, the satisfaction of agents with their partners was inferred from the perspective of the continuity of the relationship and the perception of inequities at the moment of the decision. Finally, the mutualistic benevolence of agents was derived from their satisfaction with the relationship, the value of the agreement under assessment derived from the complexity of the task, and the number of alternate relationships. A detailed description of this model is presented in [1].

Selection Decision. Every experiment had a predefined number of rounds. A different selection process was initiated by each truster at every round, by generating and announcing the complexity (contextual dimension d_6) and the deadline (d_7) of the task. The task conditions were then transmitted to all potential partners (represented by set $\mathcal{Y}$) through a call for proposals (cfp). In response, these partners proposed (randomly generated) values for the complexity and deadline of the task that were more or less close to the ones specified in the cfp. We used then a heuristic to compute the 'utility' of each proposal (up) based on the shift of the proposed values to the cfp values. Finally, each truster selected the highest rated partner based on the candidates' trustworthiness ($tw_{x,y}$) and the utility of their proposals, as follows: $\arg\max_{y_i \in \mathcal{Y}}(1/2tw_{x,y_i} + 1/2up_{y_i})$.

Types of Trusters. We considered three different basic types of trusters: B agents, which used the well know Beta Reputation trust-based evidence aggregation algorithm [2] to compute the trustworthiness scores; J agents, which used the well-known asymmetry-based trust update function defined in [3] to estimate trustworthiness; and S agents, which used our aggregator *Sinalpha* ([4]).

6.2 Experiments and Results

In this set of experiments, we wanted to test Hypothesis 1, which we reformulated as follows: *In the presence of populations of trusters and trustees that evolve their behavior based on the benevolent relationships they are able to develop with each others, trusters that are able to extract the benevolence of the trustees toward the trusters from the available evidence using* Social Tuner *will perform better than those that do not have this ability.*

Then, we defined three new types of agents, BB, JB and SB, which used a trustworthiness evaluation function composed of the basic trust-based evidence aggregator (used in B, J and S, respectively) combined with the functionalities of *Social Tuner*, as defined in in Algorithm 1. Hence, we are evaluating the benefits of using *Social-Tuner* when applied to different types of trustworthiness estimators. These last trusters performed an additional selection procedure, described as follows: just before ordering the proposals by trustworthiness and utility, each truster removed from the set of all considered proposals these proposals owned by trustees that did not reach a benevolence threshold given by the average of the mean and the maximum benevolence values presented by all candidates. Hence, we ran six different types of trusters simultaneously (B, J, S, BB, JB and SB), each with four agents.

In order to compare all approaches, we measured and averaged the number of agreements with outcomes F, Fd and V, as well as the utility of the proposals of the selected trustees. Moreover, in order to better evaluate the effect of using the *Social Tuner* component in different conditions regarding the number of interactions between trusters and trustees, we further ran the experiments with 20, 50, and 100 rounds. We set $N_{ben_{close}} = 15$ (cf. Algorithm 1).

The results of these experiments in terms of outcomes of type F and V (including mean M and standard deviation SD) are systematized in Table 1. We verified that the effect of adding *Social Tuner* to a simple trust-based evidence aggregator depended on the number of rounds considered. For instance, with only 20 rounds, when the number of interactions between any two partners was not large, there was an increase in the number of outcomes of type F for all trusters using *Social Tuner* when compared with their benevolence-less counterparts, but the difference was not statistically significant when using t-tests with Bonferroni adjustments (SB/S: +2.20% of outcomes of type F, $t(29) = 1.34$, $p = 0.09$; BB/B: +4.72% F, $t(29) = 2.21$, $p = 0.02$;JB/J: +1.09% F, $t(29) = 0.66$, $p = 0.26$). With 50 rounds, we verified that the addition of *Social Tuner* increased the number of outcomes of type F at least in 5.38%, for all basic aggregators, with all results being statistically significant (SB/S: +5.89% F, $t(29) = 2.85$, $p < 0.008$; BB/B: +6.05% F, $t(29) = 3.87$, $p < 0.008$;JB/J: +5.38% F, $t(29) = 4.61$, $p < 0.008$). Finally, with 100 rounds, we got an even more relevant increase in the number of outcomes F using *Social Tuner*, confirming that this component is in fact effective in capturing the benevolence existing between any pair of trusters-trustees(SB/S: +9.79% F, $t(29) = 5.41$, $p < 0.008$; BB/B: +7.52% F, $t(29) = 5.55$, $p < 0.008$;JB/J: +8.41% F, $t(29) = 6.60$, $p < 0.008$). The results obtained concerning outcomes of type V where in line of those just described, as can be confirmed from Table 1.

Table 1. Outcomes of types F and V per truster type and number of rounds

	outcome F						outcome V					
	20 rounds		50 rounds		100 rounds		20 rounds		50 rounds		100 rounds	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
S	0.817	0.081	0.803	0.113	0.770	0.105	0.093	0.052	0.110	0.066	0.119	0.066
SB	0.835	0.067	0.850	0.095	0.845	0.091	0.077	0.047	0.080	0.063	0.078	0.057
B	0.762	0.090	0.777	0.074	0.753	0.069	0.133	0.070	0.132	0.052	0.142	0.050
BB	0.798	0.084	0.824	0.091	0.810	0.070	0.107	0.055	0.096	0.056	0.101	0.041
J	0.827	0.084	0.799	0.075	0.765	0.067	0.088	0.052	0.115	0.050	0.129	0.043
JB	0.836	0.071	0.842	0.081	0.830	0.081	0.076	0.046	0.078	0.044	0.088	0.046

We intended to further test Hypothesis 1, and then we formulated an additional hypothesis, described as follows: *In the presence of populations of trusters and trustees of homogeneous benevolence, trusters that use the* Social Tuner *component will perform no worse than those that do not use this component.*

In order to test this new hypothesis, we made important changes to the behavioral model of agents described before. First, we set the dispositional benevolence of both trusters and trustees to a fixed value of *Medium.* Second, the ability in agreement, which determines whether the trustees fulfill or delay their agreements given the effort required to perform the agreement, is no longer dependent on the benevolence of these trustees toward the exchange partner, and is given solely by the trustees' ability (for more on this, cf. [1]). Hence, the resulting agents are not driven by benevolence.

We ran this set of experiments again with six different types of trusters running simultaneously, each with four agents: S, SB, B, BB, J, and JB. All experiments had 100 rounds. The results of these experiments, in terms of outcomes of types F and V, are systematized in Table 2.

Table 2. Outcomes of type F and V per truster type (100 rounds)

	outcome F		outcome V	
	M	SD	M	SD
S	0.941	0.055	0.028	0.028
SB	0.947	0.061	0.025	0.030
B	0.910	0.058	0.048	0.034
BB	0.933	0.059	0.031	0.028
J	0.927	0.056	0.036	0.028
JB	0.938	0.054	0.027	0.026

From the results, we observed that no one of the three chosen trust-based evidence aggregators (i.e., S, B, and J) performed more poorly when combined with the *Social Tuner* component. In fact, all of them performed a little better in terms of outcome F, although this increase was only statistically significant (using t-tests and Bonferroni adjustments) with model B ($t(29) = -4.51, p < 0.003$). The same happened when measuring the outcomes of type V, where the decrease observed with model B was statistically significant ($t(29) = 5.67, p < 0.003$). Overall, in the conditions of these two sets of experiments, we were able to confirm the truthfulness of Hypothesis 1.

7 Discussion and Conclusions

Computational trust is crucial for well-based decision making regarding possible agents' future joint activities. It heavily relies on the estimation of trustworthiness to assess the trust on particular trustees. To better estimate this trustworthiness, it is important to estimate, besides other relevant features, their ability and benevolence separately, and to combine them taking into consideration the particular situation and relationship. However, the majority of the computational trust approaches presented in literature estimates the trustworthiness of agents as a block and does not consider its dimensions in an individual form. The exception is the model of [8], which, however, assumes the existence of explicit information on benevolence, and does not present any alternative mechanism to infer this benevolence from past actions, when it is needed.

In this paper, we described a part of our computational model of trust, which is a novel approach based on the socio-cognitive trust theory that produces individual estimations of the ability and benevolence of trustees and combines these estimations in a dynamic way, taking into account the relationship existing between truster and trustee at any given moment and situation. We evaluated our approach in a simulated experimental environment by comparing three known trust-based evidence aggregators with three enhanced versions of these aggregators; the benevolence-enhanced models aggregated the values of the estimated ability (as calculated by the simple aggregators) with the estimated benevolence (as calculated by *Social Tuner*) in a dynamic way, where the weight of benevolence in the trustworthiness formula grew with the increasing number of interactions between any truster-trustee pair. To perform the comparison, we measured the outcomes of the interactions between trusters and trustees with and without the addition of *Social Tuner*. Besides, we went beyond traditional evaluation of computational trust models and used a model of agents' behavior where both trusters and trustees evolve their behaviors based on personality traits, mutualistic interests and the stage of the different relationships existing between the agents.

The results have shown that, using exactly the same evidential datasets, the approaches that included the addition of *Social Tuner* increased the number of outcomes of type F and decreased the number of outcomes of type V for all of these trust-based aggregators, for all number of rounds considered. Therefore, we concluded that the use of *Social Tuner* allowed for a significant improvement in trustworthiness estimation, leading to better computational trust models in the described situations. Concerning future work, we intend to further identify the particular circumstances in which the use of this sophisticated trust model is more relevant. Also, we intend to explore integrity as another dimension of trustworthiness, as well as to explore other ways of combining the trustworthiness dimensions, and to use other antecedents of trust, such as the trusters' own propensity to trust.

References

1. Urbano, J., Rocha, A.P., Oliveira, E.: An approach to computational social trust. AI Communications (accepted for publication in February 2013)
2. Jøsang, A., Ismail, R.: The beta reputation system. In: Proc. 15th Bled Electronic Commerce Conf. (2002)
3. Bosse, T., Jonker, C.M., Treur, J., Tykhonov, D.: Formal analysis of trust dynamics in human and software agent experiments. In: Klusch, M., Hindriks, K.V., Papazoglou, M.P., Sterling, L. (eds.) CIA 2007. LNCS (LNAI), vol. 4676, pp. 343–359. Springer, Heidelberg (2007)
4. Urbano, J., Rocha, A.P., Oliveira, E.: Computing confidence values: Does trust dynamics matter? In: Lopes, L.S., Lau, N., Mariano, P., Rocha, L.M. (eds.) EPIA 2009. LNCS, vol. 5816, pp. 520–531. Springer, Heidelberg (2009)
5. Hardin, R.: Trust and trustworthiness. The Russell Sage Foundation series on trust. Russell Sage Foundation, New York (2002)
6. Kiyonari, T., Yamagishi, T., Cook, K.S., Cheshire, C.: Does trust beget trustworthiness? trust and trustworthiness in two games and two cultures: A research note. Social Psych. Q. 69(3), 270–283 (2006)
7. Colquitt, J.A., Scott, B.A., LePine, J.A.: Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. Journal of Applied Psychology 92, 909–927 (2007)
8. Castelfranchi, C., Falcone, R.: Trust Theory: A Socio-Cognitive and Computational Model. John Wiley & Sons Ltd., Chichester (2010)
9. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. The Academy of Management Review 20(3), 709–734 (1995)
10. Schoorman, F.D., Mayer, R.C., Davis, J.H.: An integrative model of organizat. trust: Past, present, and future. Ac. Manag. Rev. 32(2), 344–354 (2007)

11. Kelton, K., Fleischmann, K.R., Wallace, W.A.: Trust in digital information. J. Am. Soc. Inf. Sci. Technol. 59(3), 363–374 (2008)
12. Elangovan, A.R., Shapiro, D.L.: Betrayal of trust in organizations. Ac. Manag. Rev. 23(3), 547–566 (1998)
13. Levin, D.Z., Cross, R., Abrams, L.C., Lesser, E.L.: Trust and knowledge sharing: A critical combination. In: Lesser, E., Prusak, L. (eds.) Creating Value with Knowledge, pp. 36–43. Oxford University Press (2004)
14. Xie, Y., Peng, S.: How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness. Psychology and Marketing 26(7), 572–589 (2009)
15. Adali, S., Wallace, W.A., Qian, Y., Vijayakumar, P., Singh, M.P.: A unified framework for trust in composite networks. In: Proc.14th AAMAS W. Trust in Agent Societies, Taipei, pp. 1–12 (May 2011)
16. Lee, D.-J., Jeong, I., Lee, H.T., Sung, H.J.: Developing a model of reciprocity in the importer exporter relationship. Indust. Market. Management 37(1), 9–22 (2008)
17. Platek, S.M., Krill, A.L., Wilson, B.: Implicit trustworthiness ratings of self-resembling faces activate brain centers involved in reward. Neuropsychologia 47(1), 289–293 (2009)
18. Koscik, T.R., Tranel, D.: The human amygdala is necessary for developing and expressing normal interpersonal trust. Neuropsyc. 49(4), 602–611 (2011)
19. McCullough, M.E., Hoyt, W.T.: Transgression-related motivational dispositions. Pers. Social Psych. Bull. 28(11), 1556–1573 (2002)
20. Roccas, S., Sagiv, L., Schwartz, S.H., Knafo, A.: The big five personality factors and personal values. Personal. Soc. Psychol. Bull. 28(6), 789–801 (2002)
21. Bouchard, T.J., McGue, M.: Genetic and environmental influences on human psychological differences. J. Neurobiology 54(1), 4–45 (2003)
22. Srivastava, S., John, O.P., Gosling, S.D., Potter, J.: Development of personality in early and middle adulthood: Set like plaster or persistent change? J. Personal. Soc. Psychol. 84(5), 1041–1053 (2003)
23. Allison, P.D.: The cultural evolution of beneficent norms. Social Forces 71(2), 279–301 (1992)
24. Foddy, M., Platow, M.J., Yamagishi, T.: Group-based trust in strangers. Psychological Science 20(4), 419–422 (2009)
25. Lewis, J.D., Weigert, A.: Trust as a social reality. Social Forces 63(4), 967–985 (1985)
26. Hardin, R.: Trust and society. In: Galeotty, G., Slamon, P., Wintrobe, R. (eds.) Competition and Structure, pp. 17–46. Cambridge Univ. Press (2000)
27. Sako, M.: Does trust improve business performance? In: Lane, C., Bachmann, R. (eds.) Trust within and between Organizations. Oxford University Press (1998)
28. Ireland, R.D., Webb, J.W.: A multi-theoretic perspective on trust and power in strategic supply chains. J. Op. Management 25(2), 482–497 (2007)
29. Venanzi, M., Piunti, M., Falcone, R., Castelfranchi, C.: Facing openness with socio-cognitive trust and categories. In: IJCAI, pp. 400–405 (2011)
30. Herzig, A., Lorini, E., Hübner, J.F., Vercouter, L.: A logic of trust and reputation. Logic J. IGPL 18(1), 214–244 (2010)
31. Abowd, G.D., Dey, A.K., Brown, P., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 304–307. Springer, Heidelberg (1999)
32. Huynh, T.D., Jennings, N.R., Shadbolt, N.R.: An integrated trust and reputation model for open multi-agent systems. Autonomous Agents and Multi-Agent Systems 13, 119–154 (2006)