

Trust and Normative Control in Multi-Agent Systems: An Empirical Study

Joana Urbano, Henrique Lopes Cardoso, Ana Paula Rocha, Eugénio Oliveira

Abstract Despite relevant insights from socio-economics, little research in multi-agent systems has addressed the interconnections between trust and normative notions such as contracts and sanctions. Focusing our attention on scenarios of betrayal, in this paper we combine the use of trust and sanctions in a negotiation process. We describe a scenario of dyadic relationships between truster agents, which make use of trust and/or sanctions, and trustees characterized by their ability and integrity, which may influence their attitude toward betrayal. Both agent behavior models are inspired in socio-economics literature. Through simulation, we show the virtues and shortcomings of exploiting trust, sanctions and a combination of both.

1 Introduction

Computational trust models are important when addressing social relationships, both in the real world and in artificial agent societies. Trust is used to tackle the problem of social control. Additionally, norms and sanctions also play an important part by exerting pressure on individuals to conform. A number of studies from the social sciences look into the interrelationships between these issues.

Social relations are often associated with uncertainty and vulnerability of interacting partners, especially while no trust relationship is formed yet. This eventually leads to opportunism, which can be defined as “some form of cheating or undersupply relative to an implicit or explicit contract” [11]. Governance mechanisms used to reduce opportunism include control and monitoring. In fact, one approach to break out low trust dynamics in bilateral relationships is to use legalistic remedies, including the use of formal contracts [7]. Luhmann [5] suggests that the existence of legal norms is one of the most effective remedies to confine the risk associated with lack

Joana Urbano, Henrique Lopes Cardoso, Ana Paula Rocha and Eugénio Oliveira
LIACC / DEI, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465
Porto, Portugal, e-mail: {joana.urbano, hlc, arocha, eco}@fe.up.pt

of trust, making it more comfortable for a potential truster to decide to invest trust in a relationship: legal regulations and sanctions reduce the risk of being betrayed. Furthermore, legal norms can foster the constitution of trust [1]. However, designing detailed contracts involves substantial drafting costs [11, 12], as does monitoring. According to Mayer *et al.*, legalistic remedies may bring organizational legitimacy but are often ineffective, being described as “impersonal substitutes for trust” [6].

At this point, it seems reasonable to think that trust can be used to reduce transaction costs associated with negotiation, monitoring and enforcement [4]. On the other hand, monitoring and enforcement have an important role when there is not appropriate evidence on which trust can be built, making the act of trust a risky engagement [5]. These concepts are therefore interconnected.

Das and Teng [2] state that trust denotes expectancies about other agents’ motivations; it is not meant to influence or affect their behavior. Control mechanisms, on the other hand, may be used with the intent of deterring opportunism. Das and Teng distinguish trust in a partner from confidence in a transaction, that is, the certainty about cooperative behavior regardless of the possible motivations of the trustee. Accordingly, confidence in a transaction may be obtained as a combination of trust and control: for the same level of confidence, if we trust less, we use more control mechanisms. Furthermore, trust and control are seen as parallel and supplementary notions: they contribute independently to the level of confidence. Any one of these mechanisms may be used if an increase in transaction confidence is needed.

Despite these relevant insights from socio-economics, little research in multi-agent systems has addressed the interconnections between trust and normative notions such as contracts and sanctions. In this paper we try to cover this gap by combining, in a negotiation process, the use of trust and contractual sanctions in scenarios of betrayal, where violations of commitments are voluntary and harmful for the betrayed entity. More specifically, we formalize a model of agents (inspired in socio-economics literature) for dyadic relationships between truster agents that make use of trust and/or sanctions as a control mechanism, and trusted agents (or trustees) that are characterized by their ability and integrity, which dictate their bias toward betrayal. We demonstrate in the paper that several forces must be weighted in the interrelation between trust and norms, such as the ability and integrity of the trustees, the sanctioning costs and the motivation to betray, and that the quantification of these forces is not a trivial task.

This paper is structured as follows. Section 2 describes the scenario underlying our study and presents the behavior models for both trusters and trustees. Section 3 describes a set of experiments we have performed and interprets the results obtained. Finally, Section 4 concludes the paper.

2 Scenario and Agents Behavior Model

We here describe our interaction scenario and present the behavior model of the agents considered in this study. We address dyadic relationships with clients and

providers of services. Clients are trusting agents (i.e. the *trustors*) who need to select the best providers for interaction, and providers are trusted agents (i.e. the *trustees*).

A trustor starts by sending a call-for-proposals for a particular service, for which each trustee will provide its own proposal. When assessing proposals, trustors take into account their utility and (optionally) the perceived trustworthiness of each proponent. The trustor will try to establish a contract with the proponent of the better assessed proposal, for which it may decide to include control mechanisms in the contract. If, for some reason, the trustee is not able to accommodate this contract, the trustor will try with the proponent of the second best proposal, and so on. At the contract enactment phase, each hired trustee will have the opportunity to fulfill the contract or to violate it, according to the behavior model described later.

2.1 Trustors

The trustor behavior model is based on the interplay between trust and control, as discussed in [2] and [8]. When considering the establishment of a contract with a trustee, the trustor computes a confidence threshold C_t that indicates the minimum confidence he needs for entering into that particular transaction. This value is calculated by weighting the perceived risk R by the agent's risk aversion Ra . Risk, in turn, is modeled as a function of the weight of the transaction volume Tv on the agent's overall production volume Pv and the perceived trustworthiness T of the trustee, computed dynamically using a computational trust model. We thus have that $C_t = R * Ra$, where risk $R = Tv/Pv * (1 - T)$.¹ Risk aversion ranges from 0 (a risk lover agent) to 1 (totally risk averse).

Having a minimum confidence threshold, the trustor will propose, to a selected trustee, a contract that includes a level of control (represented as a sanction to apply in case of violation) computed according to the general notion from [2] that $Confidence = Trust + Control$. By suggesting an appropriate sanction, the trustor tries to raise his confidence on the contract that is to be established with a particular trustee, of which it has some trustworthiness assessment.

2.2 Trustees

The model of behavior of the trustees is inspired in the model of betrayal in organizations of Elangovan and Shapiro [3]. In our model, trustees of low integrity tend to enter in new contracts even when they do not have enough resources to satisfy them, i.e. they are aware that they may have to (voluntarily) violate one or more of their

¹ In the experiments, we use T/ζ instead of T , due to the fact that computational trust models typically overrate the trustworthiness estimations, as they tend to aggregate the outcomes of past evidence using statistical methods, without taking into consideration the *relationship* that was active between interacting partners at the evidence time [10].

active contracts. On the contrary, trustees with high integrity may refuse the contract if they do not have enough resources to satisfy the deal without violating previous agreements. We also assume that all trustees have a predefined level of *competence*, i.e. an innate ability to provide products of good quality. Violations that are due to (lack of) ability are not voluntary, and thus are not considered *betrayals* [3].

In the equations that follow x denotes a trustee, y denotes a truster, c denotes a contract, and p denotes a contract proposal. When time-stamping terms using a superscript, we assume a discrete time line. Unless otherwise noted, variables are assumed to be universally quantified.

The decision to betray vs. keep the status quo is made when the trustee is awarded a new contract for which it does not possess enough resources (e.g. a stock that is periodically replenished). The new contract is considered a *business opportunity* if it presents significant higher utility than at least one of the trustee's ongoing contracts; this assessment also depends on the integrity of the trustee, as shown in Equation 1, where $\delta \in [0, 1]$ is the integrity parameter.

$$\begin{aligned} \exists c \text{ Contract}(c,x,y)^t \wedge \text{ContractProposal}(p,y,x)^t \wedge \text{Utility}(c,x) < \text{Utility}(p,x) * (1 - \delta) \\ \wedge \text{Resources}(p,x) > \text{FreeResources}(x)^t \Rightarrow \text{NewOpportunity}(p,x) \end{aligned} \quad (1)$$

After identifying a new opportunity, the trustee is going to assess the current situation, namely: i) the benefits of betraying; and ii) its relationship with the potential victim of betrayal.

Assessing the Value of Betraying. The trustee assesses the benefits of betraying by taking into account both the utility associated with the new opportunity and the existence of a relevant sanction associated with the potential contract to betray. This sanction is considered *irrelevant* to the trustee if its value is smaller than a given (adjustable) percentage γ of the utility associated with the new opportunity. In this case, the value of betrayal is high. In order to reduce the complexity of the model, we chose three qualitative values for the value of betraying, as illustrated in Equation 2.

$$\begin{aligned} \text{Utility}(p,x) - \text{Sanction}(c,x) < \gamma_1 * \text{Utility}(c,x) \Rightarrow \text{VBetrayal}(c,x, \text{low}) \\ \gamma_1 * \text{Utility}(c,x) \leq \text{Utility}(p,x) - \text{Sanction}(c,x) < \gamma_2 * \text{Utility}(c,x) \Rightarrow \text{VBetrayal}(c,x, \text{medium}) \\ \gamma_2 * \text{Utility}(c,x) \leq \text{Utility}(p,x) - \text{Sanction}(c,x) \Rightarrow \text{VBetrayal}(c,x, \text{high}) \end{aligned} \quad (2)$$

Assessing the Value of the Relationship. The trustee assesses the relationship with the potential victim by considering the number of past contracts between both partners in the last σ units of time (perspective of continuing the relationship, cf. Equation 3 for $\sigma = 3$), and the existence of other contracts (if the trustee is currently engaged in at least ξ contracts with other trusters, cf. Equation 4 for $\xi = 2$). The perceived value of the relationship is given in Equation 5.

$$\begin{aligned} \exists c_1, c_2, c_3 \text{ Contract}(c_1,x,y)^{t-1} \wedge \text{Contract}(c_2,x,y)^{t-2} \wedge \text{Contract}(c_3,x,y)^{t-3} \\ \Rightarrow \text{PerspContinuity}(x,y, \text{high})^t \end{aligned} \quad (3)$$

$$\begin{aligned} \exists c_1, c_2 \text{ Contract}(c,x,y)^t \wedge \text{Contract}(c_1,x,y_1)^t \wedge \text{Contract}(c_2,x,y_2)^t \wedge \\ c_1 \neq c_2 \wedge y_1 \neq y \wedge y_2 \neq y \Rightarrow \text{HasOtherContracts}(x)^t \end{aligned} \quad (4)$$

$$\begin{aligned}
& \neg \text{PerspContinuity}(x,y,\text{high})^t \wedge \text{HasOtherContracts}(x)^t \Rightarrow \text{VRelationship}(x,y,\text{low})^t \\
& \qquad \qquad \qquad \neg \text{PerspContinuity}(x,y,\text{high})^t \wedge \neg \text{HasOtherContracts}(x)^t \vee \\
& \text{PerspContinuity}(x,y,\text{high})^t \wedge \text{HasOtherContracts}(x)^t \Rightarrow \text{VRelationship}(x,y,\text{medium})^t \quad (5) \\
& \text{PerspContinuity}(x,y,\text{high})^t \wedge \neg \text{HasOtherContracts}(x)^t \Rightarrow \text{VRelationship}(x,y,\text{high})^t
\end{aligned}$$

The decision to betray a partner (accepting the new contract) or instead to keep its trust takes into consideration the assessment made by the trustee concerning the values of betrayal and relationship. In case there is more than one contract which is deemed to be betrayed, the trustee will only betray the one with less utility, provided that its allocated resources are enough to take into account the new contract. In no contract is deemed to be betrayed, the trustee declines to accept the new contract.

It is important to note that even new contracts may be betrayed later on if another opportunity arises. Contracts are violated at enactment time, which means that the decision to betray is made much earlier than the *act* of betray.

3 Experiments

In this section, we present the experiments we have run as support to our study. We used the trading scenario described in detail in the previous section.

3.1 Experimental Setup

We ran all the experiments using the computational trust model described in [9]. Each experiment was composed of 80 rounds, and at every round each buyer started a new negotiation cycle by issuing a new call for proposals. At the first round of each experiment, the repository of trust evidence for every supplier was cleaned. We used 80 buyers and 120 suppliers. Every experiment was run 30 times. Table 1 shows the values we assigned to the model's variables in the experiments.

Table 1 Configuration of parameters

| | δ_{low} | δ_{medium} | δ_{high} | γ_1 | γ_2 | σ | ξ | ζ |
|-------|----------------|-------------------|-----------------|------------|------------|----------|-------|---------|
| value | 0.3 | 0.6 | 0.9 | 0.3 | 0.9 | 3 | 1 | 4 |

The effective betrayal of contracts was configured probabilistically taking into consideration the assessed values of the benefits of betraying and of the relationship. This probability was 1.0 for high benefits and low and medium relationship values; for high benefits and high relationship values, or medium benefits and low relationship values, the probability of betrayal was 0.5; finally, a betrayal happened

with probability 0.2 when both the values of the benefits of betraying and of the relationship were medium. In all the remaining cases, the trustees did not betray the trusters.

Configuration of Trusters. The sanction value was calculated as $S = Ct - T/\zeta$. This formula provides the relationship between the trustworthiness of a trustee and the level of sanctions S that a truster will propose to be included in the contract. We start from the formulation of $Ct = T + S$, where for the reason explained before we reduce the weight of the trust parcel. Every truster has a value $Ra \in [0, 1]$ picked randomly at setup, and a value Pv also picked up randomly from a range of fixed minimum and maximum values. Tv is a dynamic value proposed by a trustee resulting from a specific contract negotiation.

After a betrayal, the truster enters a resent period toward the offender trustee, meaning that the trustworthiness of the latter drops to zero immediately after the betrayal and is gradually softened with time. This effect on the trustee's trustworthiness T is given by $T = T * \frac{\Delta_t}{\rho}$, where Δ_t represents the elapsed time (in these experiments, $\rho = 3$ rounds) since the time of the offense.

The population of trusters follows an uniform distribution over the possible types:

| | |
|-----------|--|
| None | The truster does not use sanctions nor trust. |
| Sanctions | The truster use sanctions but does not select partners based on the trust. |
| Trust | The truster uses trust to select partners but does not use sanctions. |
| S&T | The truster uses trust both to select partners and to compute sanctions. |

Configuration of Trustees. In order to emulate the existence of a potential new opportunity (cf. Equation 1), all suppliers had a limited stock within a simulation round. The utility of a contract for a trustee is calculated by multiplying the dimension of the proposal (i.e. the quantity of material agreed over the available stock of the trustee) by the relevance of the price in the proposal.

The innate ability of trustees was configured as a random value ascribed at setup in range $[0.5, 1]$, and their integrity were also randomly picked up considering the possible δ values (Table 1).

Evaluation Metrics. In these experiments, we used eight different performance metrics: Δ_{sup} (no. of different suppliers selected by all buyers in one round); τ , α and ι (average trustworthiness, ability and integrity of the suppliers selected by the buyers in one round, respectively); o^+ (number of contracts with positive outcome in a round); Σ (average sanction applied by all buyers to the contracts they establish with the suppliers, in one round); O and β (opportunities to betray faced by the suppliers, and the effective betrayals occurring in one round, respectively). All metrics took values in $[0, 1]$, all averaged over all rounds and all runs of the experiments.

3.2 Results

Experimental results are shown in Table 2. From the table, we verify that there is a clear difference in results when we consider the use of trust in negotiation and when we do not. In fact, approaches that do not use trust (None and Sanctions) select the partners based only on the utility of their proposals. Consequently, there is a wider choice of suppliers (Δ_{sup}) and the average values of trustworthiness (τ) and ability (α) of the selected suppliers is rather low. Also, as more suppliers are selected per round, the less the chance that they overpass their stock, and the opportunities of betrayal (O) are, thus, smaller than in the approaches that use trust (Trust and S&T). In contrast, the latter are more exposed to effective betrayals (β), due to the just mentioned effect.

Table 2 The results of the experiments

| | Δ_{sup} | τ | α | ι | o^+ | Σ | O | β |
|-----------|----------------|--------|----------|---------|-------|----------|------|---------|
| None | 0.92 | 0.27 | 0.73 | 0.42 | 0.66 | 0.00 | 0.17 | 0.09 |
| Sanctions | 0.93 | 0.27 | 0.73 | 0.42 | 0.69 | 0.38 | 0.18 | 0.05 |
| Trust | 0.77 | 0.82 | 0.88 | 0.43 | 0.74 | 0.00 | 0.34 | 0.16 |
| S&T | 0.80 | 0.81 | 0.88 | 0.43 | 0.76 | 0.18 | 0.40 | 0.13 |

The most effective approach to prevent betrayals is to use sanctions standalone, although this has the disadvantage of not caring about the ability dimension of trustees, thus decreasing o^+ . Comparing approaches S&T and Trust, we verified that although the former raised the number of betrayal opportunities, the addition of sanctions allowed to reduce the number of effective betrayals in almost 20%.

3.3 Discussion

The experiments we have run allowed to shed some light on the complex interrelation between trust and sanctions. We observed that sanctions are more effective in helping on preventing betrayals, and that trust is more effective in distinguishing between entities with different abilities. However, our study shows that in a realistic world, where trustees have different ability and integrity values, the interrelation between both governance mechanisms is less evident. The supplementary use of trust and sanctions seems to be the more balanced approach in terms of contracts with positive outcome (o^+), at the expense of a moderate usage of sanctions ($\Sigma = 0.18$) and a (not irrelevant) rate of betrayals ($\beta = 0.13$). Finally, the results were not conclusive about the best governance mechanisms to prevent selection of trustees with low values of integrity.

4 Conclusions

In this paper, we presented an empirical study about the interrelations between trust and norms. This study was grounded on solid theory from diverse research areas concerning trust and norms. The novelty of this study concerns the experimental exploration of the complex relationships between factors such the ability and integrity dimensions of trustworthiness, risk, sanctions and betrayals. From the gained experience in this work, we conclude that the conjunctive use of trust and norms – described in theoretic works as a promising interrelated governance mechanism – is not a trivial task.

For future work, we intend to refine our approach to get a better insight on the problem and reach more significant results.

Acknowledgements This research is supported by Fundação para a Ciência e a Tecnologia (FCT), under project PTDC/EIA-EIA/104420/2008. The first author is supported by FCT under grant SFRH/BD/39070/2007.

References

1. Bachmann, R.: Trust, power and control in trans-organizational relations. *Organization Studies* **22**(2), 341–369 (2001)
2. Das, T.K., Teng, B.: Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review* **23**(3), 491–512 (1998)
3. Elangovan, A.R., Shapiro, D.L.: Betrayal of trust in organizations. *The Academy of Management Review* **23**(3), 547–566 (1998)
4. Ireland, R.D., Webb, J.W.: A multi-theoretic perspective on trust and power in strategic supply chains. *Journal of Operations Management* **25**(2), 482 – 497 (2007)
5. Luhmann, N.: *Trust and Power*. John Wiley & Sons, New York (1979)
6. 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)
7. Sako, M.: Does trust improve business performance? In: C. Lane, R. Bachmann (eds.) *Trust within and between Organizations: Conceptual Issues and Empirical Applications*. Oxford University Press (1998)
8. Tan, Y.H., Thoen, W.: An Outline of a Trust Model for Electronic Commerce. *Applied Artificial Intelligence* **14**(8), 849–862 (2000)
9. Urbano, J., Rocha, A.P., Oliveira, E.: Computing confidence values: Does trust dynamics matter? In: *Proceedings of the 14th Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence, EPIA '09*, pp. 520–531. Springer-Verlag, Berlin, Heidelberg (2009)
10. Urbano, J., Rocha, A.P., Oliveira, E.: A dynamic agents behavior model for computational trust. In: L. Antunes, H. Pinto (eds.) *Progress in Artificial Intelligence, LNCS*, vol. 7026, pp. 536–550. Springer Berlin / Heidelberg, Berlin, Heidelberg (2011)
11. Wathne, K.H., Heide, J.B.: Opportunism in interfirm relationships: Forms, outcomes, and solutions. *The Journal of Marketing* **64**(4), 36–51 (2000)
12. Williamson, O.E.: Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics* **22**, 233–261 (1979)