



Trust and Normative Control in Multi-Agent Systems

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

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

KEYWORD

Multi-agent systems
Computational Trust
Normative Control
Betrayal

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 trustee agents, 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 using trust, sanctions, and a combination of both in processes of selection of partners.

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 interrelationships between these issues.

Social relations are often associated with uncertainty and vulnerability of interacting partners. These relations are traditionally secured by ongoing interactions, where interacting partners trust each other [HARDIN, 2001]. In fact, social scientists often consider that trust is the fuel of society [LUHMANN, 1979; LEWIS and WEIGERT, 1985]. However, not all relations are trust relations: the reality of present days indicates the urge for new forms of relationships, mainly in business and in social networks, to be formed more quickly and, more and more, with

anonymous others, or strangers. This may happen by need, or just because opportunities for better deals exist outside the committed relation [YAMAGISHI and YAMAGISHI, 1994; MACY and SKVORETZ, 1998].

Without valid alternatives to trust, opportunism – which can be defined as “some form of cheating or undersupply relative to an implicit or explicit contract” [WATHNE and HEIDE, 2000] – will arise. Governance mechanisms used to reduce opportunism include incentives, 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 [SAKO, 1998]. Luhmann [LUHMANN, 1979] suggests that the existence of legal norms is one of the most effective remedies to confine the risk associated with lack 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 [BACHMANN, 2001]. However, designing detailed contracts involves substantial



drafting and monitoring costs, especially when monitoring is difficult or when sanctions require extensive litigation [WILLIAMSON, 1979; MAYER *et al.*, 1995; MACY and SKVORETZ, 1998; WATHNE and HEIDE, 2000; HARDIN, 2004].

According to Mayer *et al.*, legalistic remedies may bring organizational legitimacy but are often ineffective, being described as “impersonal substitutes for trust” [MAYER *et al.*, 1995].

At this point, it seems reasonable to think that trust can be used to reduce transaction costs associated with negotiation, monitoring and enforcement [IRELAND and WEBB, 2007]. 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 [LUHMANN, 1979]. These concepts are therefore interconnected.

Das and Teng [DAS and TENG, 1998] 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 sought 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 increased 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.¹ 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 presents the results obtained. Section 4 discusses the results. Finally, Section 5 concludes the paper.

2 Scenario and agents’ behavior models

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 *trusters*) who need to select the best providers for interaction, and providers are trusted agents (i.e. the *trustees*).

A truster starts by sending a call-for-proposals (CFP) for a particular service, for which each trustee will provide his own proposal. When assessing proposals, trusters take into account their utility² and (optionally) the perceived trustworthiness of each proponent. The truster 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 truster 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 The utility of a proposal reflects how close or how far its terms are from the conditions established in the CFP. Each proposal is generated after the CFP is received so that its terms have a slight and random deviation from the CFP’s conditions.

¹ This paper is based on and extends [URBANO *et al.*, 2012].



2.1 Trusters

The truster behavior model is based on the interplay between trust and control, as discussed in [DAS and TENG, 1998] and [TAN and THOEN, 2000]. When considering the establishment of a contract with a trustee, the truster computes a confidence threshold Ct that indicates the minimum confidence he needs for entering into that particular transaction. This value is calculated by weighting the perceived risk R with 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 $Ct = R * Ra$, where risk $R = Tv / Pv * (1 - T)$.³ Risk aversion ranges from 0 (a risk lover agent) to 1 (totally risk averse).

Having a minimum confidence threshold, the truster 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 [DAS and TENG, 1998] that $Confidence = Trust + Control$. By suggesting an appropriate sanction, the truster 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 mainly inspired in the model of betrayal in organizations of Elangovan and Shapiro [ELANGO VAN and SHAPIRO, 1998]. Our model considers that each trustee has an inherent and predefined ability to provide a given product and a dispositional integrity, and that both dimensions shape the trustee's trustworthy behavior [Mayer *et al.*, 1995; ELANGO VAN and SHAPIRO, 1998; COLQUITT *et al.*, 2007].

Besides, our model considers that trustees can have two different types of faulty behavior. On the one hand, they can violate contracts due to their lack of ability or competence in providing a given product under the contractual conditions. On the other hand, they can engage in voluntary actions that severely violate the pivotal expectations of the truster. These violations take the form of *betrayal* and their consequences on trust are devastating, because they disrupt ongoing and meaningful relationships in which partners have invested material and intense emotional resources [LEWIS and WEIGERT, 1985; ELANGO VAN and SHAPIRO, 1998; FITNESS, 2001; CVETKOVICH *et al.*, 2002; FINKEL *et al.*, 2002; POORTINGA and PIDGEON, 2004]. These betrayals are related with the integrity of the trustee.

Elangovan and Shapiro's model postulates that there are some conditions observed by the trustee that trigger the assessment of the current situational satisfaction with the truster. If these conditions are absent, the trustee does not even consider betraying. These conditions are: need, crises, Machiavellianism, and opportunity recognition. In our model, we consider that the triggering conditions are the possible awarding of a new contract and the integrity of the trustee: in face of a more attractive business opportunity, the trustee may consider to voluntarily violate one of his active contracts (and, then, to betray the respective truster) in order to better accommodate the new contract. However, trustees of high integrity are more reluctant to consider the new opportunity, possibly rejecting the new contract and keeping the status quo with his active interaction partners, while low integrity trustees tend to consider the new opportunity more often.

Following Elangovan and Shapiro's model of betrayal, if the triggering conditions are strong enough to make the trustee assess the current situational satisfaction with the truster, then the trustee starts evaluating the benefits of betraying, the relationship with the truster, and his own principles (closely related with ethics and integrity). This potentially leads to the formation of the motivation to betray. Finally, the action of either betraying or keeping the status quo depends on the trustee's motivation to betray and on his assessment of the penalties associated to the betrayal [ELANGO VAN and SHAPIRO, 1998].

In what follows, we describe our model of betraying, which is a simplified version of Elangovan and Shapiro's model. We will give particular attention to

³ In our experiments we use T/ξ instead of T , since it is our belief that computational trust models typically overrate trustworthiness estimations: they tend to aggregate the outcomes of past evidence using statistical methods, not taking into consideration the *relationship* that was active between the partners at the time of each evidence [URBANO *et al.*, 2011].



the behavior of trustees toward voluntary violations, i.e., violations that do not result from faulty behavior of trustees due to ability issues.

Whenever necessary, we use a formalism in pseudo-logic that allows disambiguating important considerations of our model. 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.

The decision to betray vs. keep the status quo is made in a process that starts when the trustee is selected for a new contract. The new contract is considered a *business opportunity*. As such, the agent will probabilistically consider betraying one of his active contracts. This betray propensity (henceforth ρ) is inversely proportional to the integrity of the trustee, denoted as $\delta \in [0,1]$. We may say that with a probability ρ the trustee will consider betraying one of his active contracts c , provided that freeing the resources that are allocated for upholding that contract enables him to accept the new contract (i.e. when Equation 1 holds).

$$\begin{aligned} \text{FreeResources}(x) + \text{Resources}(c) \\ > \text{Resources}(p) \end{aligned} \quad (1)$$

Active contracts are analyzed in decreasing order of utility.⁴ It is worth noting that even when the trustee already has enough free resources to encompass the new contract, it will still consider betraying one of his active contracts.

After identifying a new opportunity, the trustee is going to assess the current situation, namely: i) the benefits of betraying; and ii) his relationship with the potential victim of betrayal. It is worth to note that Elangovan and Shapiro's model has a third situational assessment component, the assessment of principles. In our model, we decided to incorporate this component into the betray propensity parameter described above; i.e., the principles of the trustee (related with his integrity) are used as triggers to the assessment of the situation, and not (directly) as one dimension of this assessment.

⁴ The trustees evaluate the utility of their contracts taking into consideration the transactional volume and price associated to each of their contracts.

2.2.1 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 contract that is to be possibly betrayed. 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 Equations 2-4.

$$\begin{aligned} \text{Utility}(p, x) - \text{Sanction}(c, x) - \text{Utility}(c, x) \\ < \gamma_1 / (1 - \delta) \\ \Rightarrow \text{VBetrayal}(c, x, \text{low}) \end{aligned} \quad (2)$$

$$\begin{aligned} \gamma_1 / (1 - \delta) < \\ \text{Utility}(p, x) - \text{Sanction}(c, x) - \text{Utility}(c, x) \\ < \gamma_2 / (1 - \delta) \\ \Rightarrow \text{VBetrayal}(c, x, \text{medium}) \end{aligned} \quad (3)$$

$$\begin{aligned} \gamma_2 / (1 - \delta) < \\ \text{Utility}(p, x) - \text{Sanction}(c, x) - \text{Utility}(c, x) \\ \Rightarrow \text{VBetrayal}(c, x, \text{high}) \end{aligned} \quad (4)$$

2.2.2 Value of the Relationship

The trustee assesses the relationship with the potential victim by considering: (i) if the number of past contracts between both partners in the last σ units of time exceeds a minimum value λ (perspective of continuing the relationship, cf. Equation 5 where t denotes the current time step); and (ii) the existence of at least ξ other contracts in which the trustee is currently engaged (cf. Equation 6 where t denotes the current time step). The perceived value of the relationship is given in Table 1.

$$\begin{aligned} \sum_{i=1}^{\sigma} \text{Contract}(_, x, y)^{t-i} > \lambda \\ \Rightarrow \text{PerspContinuity}(x, y, \text{high}) \end{aligned} \quad (5)$$

$$\begin{aligned} \sum \text{Contract}(_, x, _)^t > \xi \\ \Rightarrow \text{HasOtherContracts}(x) \end{aligned} \quad (6)$$



<i>PerspContinuity</i>	<i>HasOtherContracts</i>	
	<i>yes</i>	<i>no</i>
<i>high</i>	medium	high
<i>¬high</i>	low	medium

Table 1. Value of the relationship

The decision to betray a partner or instead to keep his 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. If no contract is deemed to be betrayed, the trustee may still accept the new contract provided that enough free resources are available. It is important to note that even new contracts may be later decided upon to be betrayed 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 betraying.

Finally, while the trustee may decide not to betray a partner, he may still fail the contract if his ability is not good enough.

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 have run three different set of experiments, each one using a different population of trustees, characterized by distinct values for the ability and integrity of the agents. Table 2 summarizes the different populations of trustees.

Population	<i>ability</i>	<i>integrity</i>
Heterogeneous	uniform distribution within $[0,5,1]$	uniform distribution within $\{\delta_{low}, \delta_{medium}, \delta_{high}\}$
Low Integrity		δ_{low}
High Integrity		δ_{high}

Table 2. Trustee populations

In all experiments of the three sets, we used the

computational trust model described in [URBANO *et al.*, 2009]. 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 3 shows the values we assigned to the model's variables in the experiments.

The effective betrayal of contracts was configured probabilistically (see Table 4) taking into consideration the assessed values of the benefits of betraying and of the relationship.

δ_{low}	δ_{medium}	δ_{high}	$\rho_{\delta_{low}}$	$\rho_{\delta_{medium}}$	$\rho_{\delta_{high}}$
0.1	0.5	0.9	0.5	0.3	0.1
γ_1	γ_2	σ	λ	ξ	ζ
0.0	0.2	3	2	1	4

Table 3. Configuration of parameters

<i>ValueRelationship</i>	<i>ValueBetrayal</i>		
	<i>High</i>	<i>Medium</i>	<i>Low</i>
<i>High</i>	0.5	0.0	0.0
<i>Medium</i>	1.0	0.2	0.0
<i>Low</i>	1.0	0.5	0.0

Table 4. Betray probabilities

3.1.1 Configuration of Trusters

The sanction value was calculated as $S = Ct - T/\xi$. 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.

Whenever betrayed, the truster resents the betrayer by ignoring any information regarding his previous activity, which has the effect of dropping his trustworthiness value to 0 (as assessed by the betrayed truster).

The final desideratum of the experimental compo-



ment of our work was to evaluate the performance of the different combinations of trust and sanctions when applied to processes of selection of partners. Therefore, we defined the following types of trusters:

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

The population of trusters used in all sets of experiments followed a uniform distribution over the possible types described above.

3.1.2 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 in the contract over the stock of the trustee) by the relevance of the price in the proposal.

3.1.3 Evaluation Variables

In these experiments, we focused on observing the following evaluation variables:

- Δ_{sup} : number of different suppliers selected by all buyers in one round.
- o^+ : number of contracts with positive outcome (that were not violated or betrayed) in a round.
- O : number of opportunities to betray faced by the trustees. A trustee that has n active contracts when a new opportunity arises is confronted with n opportunities to betrayal.
- β : number of effective betrayals suffered by all trusters in one round.
- β/O : ratio of the number of effective betrayals to the number of opportunities to betray, indicating the effectiveness of the selection models in dissuading the trustees from betraying after identifying a new opportunity to betray.
- Ξ : number of contracts that were violated due to (lack of) ability. This variable is derived from the values of o^+ and β .
- Σ : average sanction applied by all trusters to the contracts they establish in each round.

All variables took values in $[0,1]$, all averaged over all rounds and all runs of the experiments.

3.2 Results

In this section, we present the results of the experiments, for each one of the considered populations of trustees.

In some cases, when the difference between the results of some variable in two different models were not too evident, we used a Paired Two Sample for Means (one-tail) t-Test in order to evaluate the statistical significance of the results' differences. We used Bonferroni adjustments considering 6 t-Test comparisons, which means that the obtained p-value must have been less than 0.008 for any one comparison to be considered significant (assuming we keep the overall experimentwise error rate to 5%).

3.2.1 Heterogeneous Population

The experimental results for the Heterogeneous population concerning the evaluation variables considered in Section 3.1.3 are shown in Table 5.

The results in the Δ_{sup} evaluation metric show that agents that use trust in the selection process were less exploratory (T: 0.880, T&S: 0.878) than agents that did not use trust (N: 0.963, S: 0.966). In the same way, these trust-based agents were significantly more exposed to betray, as shown by variable O (T: 0.307, T&S: 309), than the other agents (N: 0.127, S: 0.126).

One important variable to look at is the number of positive outcomes (o^+), i.e., the number of contracts that were neither betrayed nor violated due to ability issues. We can observe in Table 5 that agents of type T&S outperformed the remaining agents in this variable. Indeed, agents of type T&S got more positive outcomes than T agents (T: 0.833, T&S: 0.846, $t[1]=-5.78$, $p<0.001$) and significantly better performance than agents of types S (0.721) and N (0.708).

Another important variable is the rate of materialized betrayals (β/O). Both of the truster types using sanctions achieved better performances on this issue (the results for these types are not statistically significantly different: S: 0.132, S&T: 0.143, $t[1]=-0.85$, $p=0.20$). Agents of type N performed worse than T agents (N: 0.217, T: 0.180, $t[1]=3.18$, $p=0.002$). Taking into consideration the two best truster types in the β/O metric, we can see that T&S agents used a much lighter sanction value Σ (0.096) than S agents did (0.230).



In terms of effective betrayals (β), S agents performed best than all other agents: agents of type N got more betrayals than agents of type S (N: 0.028, S: 0.017, $t=5.51$, $p<0.001$), and T agents got more betrayals (0.055) than T&S agents (0.044).

Finally, it is evident that trust-based models are more efficient in preventing violations due to lack of

ability (Ξ) of trustees than the remaining models (N: 0.265, S: 0.262, T: 0.112, T&S: 0.180). As can be observed, S agents cannot do better than N agents concerning this issue.

	Δ_{sup}	o^+	O	β	β/O	Ξ	Σ
None (N)	0.963	0.708	0.127	0.028	0.217	0.265	0.000
Sanctions (S)	0.966	0.721	0.126	0.017	0.132	0.262	0.230
Trust (T)	0.880	0.833	0.307	0.055	0.180	0.112	0.000
T&S	0.878	0.846	0.309	0.044	0.143	0.180	0.096

Table 5. Experimental results for Heterogeneous trustee population

	Δ_{sup}	o^+	O	β	β/O	Ξ	Σ
None (N)	0.962	0.694	0.203	0.052	0.256	0.254	0.000
Sanctions (S)	0.961	0.709	0.207	0.034	0.167	0.257	0.219
Trust (T)	0.851	0.783	0.555	0.116	0.210	0.102	0.000
T&S	0.861	0.808	0.562	0.089	0.159	0.104	0.093

Table 6. Experimental results for Low Integrity trustee population

	Δ_{sup}	o^+	O	β	β/O	Ξ	Σ
None (N)	0.967	0.732	0.046	0.006	0.120	0.263	0.000
Sanctions (S)	0.967	0.735	0.045	0.002	0.051	0.262	0.217
Trust (T)	0.894	0.866	0.108	0.011	0.107	0.122	0.000
T&S	0.896	0.870	0.109	0.010	0.092	0.120	0.086

Table 7. Experimental results for High Integrity trustee population

3.2.2 Low Integrity Population

The experimental results for the Low Integrity populations are shown in Table 6.

Similarly to what happen with the heterogeneous population, we verified that trust-based models were less explorative (Δ_{sup}) than the remaining models (N: 0.962, S: 0.961, T: 0.851, T&S: 0.861), which is inversely correlated with the opportunities to betray (N: 0.203, S: 0.207, T: 0.555, T&S: 0.562).

In the same way, we verified once again that the models that use sanctions were more effective regarding the rate of materialized betrayals (β/O), where the mean value obtained by agents of type S (0.167) was not significantly different from the equivalent mean

value for T&S agents (0.159) ($t[1]=0.99$, $p=0.17$). Concerning these truster types, T&S agents were able to use a lighter value of sanction Σ (0.093) than S agents (0.219). Agents that use trust but not sanctions could, even though, achieve a better value for β/O than agents that did not use trust neither sanctions (N: 0.256, T: 0.210).

In terms of the total value of positive outcomes (o^+), both trust-based truster types outperformed the types that do not use trust in the selection process. T&S got more positive outcomes than T agents, benefiting from using sanctions (T: 0.783, T&S: 0.808, $t[1]=-4.85$, $p<0.001$). Concerning the trusters that do not use trust, S agents got a light advantage of about



2% over N agents (N: 0.694, S: 0.709, $t[1]=-4.04$, $p<0.001$).

Once again, it is evident that trust-based models are more efficient in preventing violations due to the lack of ability (Ξ) of trustees than the remaining models (N: 0.254, S: 0.257, T: 0.102, T&S: 0.104). The results of T agents and T&S agents were not statistically significantly different (T: 0.102, T&S: 0.104, $t[1]=-1.10$, $p=0.14$), and neither were the results of N and S agents (N: 0.254, S: 0.257, $t[1]=-0.85$, $p=0.20$).

Finally, S agents were the ones that got less betrays, in absolute terms (N: 0.052, S: 0.034, T: 0.116, T&S: 0.089).

3.2.3 High Integrity Population

The experimental results for the High Integrity population are shown in Table 7.

The results confirm the expected lower values of exploration of agents that use trust-based models when compared with the those that do not use trust to select partners (N: 0.967, S: 0.967, T: 0.894, T&S: 0.896). T and T&S agents' exploration rate is about 92%-93% of the rate of N and S agents, a value that is greater to what happened when using the population Low Integrity (88%-90%). This may be explained by the fact that the population in general is more trustworthy (as the integrity of agents is higher and the ability remained unchanged) and therefore trust-based agents are less parochial in their selection choices.

The values of positive outcomes obtained using trust-based models were 18-19% higher than those obtained by trusters that do not use trust. The mean values of $o+$ obtained by both T and T&S agents were not significantly different (T: 0.866, T&S: 0.870, $t[1]=-1.55$, $p=0.067$). This may indicate that when the opportunities to betrayal are low, the use of sanctions may not be relevant. The mean values of $o+$ obtained by N agents and S agents were not significantly different (N: 0.732, S: 0.735, $t[1]=-1.30$, $p=0.10$).

As the population shows high integrity, the opportunities to betray are rather low for all types of agents (N: 0.046, S: 0.045, T: 0.108, T&S: 0.109), which translates also in low values of effective betrayals. We verified that N agents suffered more betrayals than S agents (N: 0.006, S: 0.002, $t[1]=2.76$, $p=0.005$), and that the mean values of betrayals obtained by both T and T&S agents were not signifi-

cantly different (T: 0.011, S&T: 0.010, $t[1]=2.11$, $p=0.021$). In any case, all these values are almost residual, even for trust-based models.

Once again, we verified that the trusters that use sanctions were more effective regarding the rate of materialized betrayals (β/O), where the mean value obtained by S agents was not significantly different from the equivalent mean value for T&S agents (S: 0.051, S&T: 0.092, $t[1]=-2.20$, $p=0.018$). T&S agents were able to use a lighter value of sanction (0.086) than S agents (0.217).

Finally, it is evident that trust-based models were more efficient in preventing violations due to the lack of ability (Ξ) of trustees than the remaining models (N: 0.263, S: 0.262, T: 0.122, S&T: 0.120).

4 Discussion

The results presented in the previous section for the different populations have shown that T&S is the best selection model concerning the total number of positive outcomes, i.e., the number of contracts that were neither failed due to lack of ability of trustees nor betrayed by them. This is due to the combined effect of using trust – which proved to be very effective in avoiding trustees with lower ability – and sanctions – which have shown to have an important role in persuading the trustees to maintain the status quo after identifying an opportunity to betray.

In our study, we settle an extremely complex scenario, with different models of behavior for trusters and trustees, where both models were inspired in theoretical works on normative control and trust. However, the novelty of such an approach came with a price: the resulting model had a great number of variables, potentially influencing each other, which hardened the analysis of the experimental results. For instance, a not irrelevant bias of our model is related to the fact that trusters that selected partners based on utility and trust were more exposed to potential betrayals (as shown by the opportunities of betrayal variable) than trusters that selected based only on the utility of the proposals. Indeed, the fact that the latter explored more partners implied that each trustee had fewer active contracts at one time, and consequently less potential contracts to consider betraying when a new opportunity arose.

The mentioned bias reflected directly on the results of the betrayal variable. Therefore, it is probably wiser to take into consideration the results of the β/O



variable (i.e., how many opportunities of betrayal do materialize into an effective betrayal) rather than the results on the betrayal variable, in order to understand the effectiveness of each selection model in preventing betrayals. We verified that both selection models that use sanctions (S and T&S) have shown similar performance in this variable. However, the T&S model had the additional advantage of using lighter sanctions. This happened, once again, due to the complementary action of trust and sanctions: by selecting the most trustworthy agents and considering that sanctions were drafted (also) taking into account the perceived trustworthiness of trustees, the value of the applied sanction was reduced.

The study we presented in this paper is novel, in the sense that it experimentally analyzed the combined effect of trust and normative control (in the form of sanctions) in the process of selecting partners. Also, it used models of behavior more realistic than the models generally used to test trust and norms (standalone), which are generally probabilistic and static. However, our model and our work present limitations that must be taken into consideration in future work. For instance, even though the behavioral models drink from theoretical insights on trust and sanctions, the overall model is not empirically grounded.

A second limitation concerns the simplifications that were done to the Elangovan and Shapiro's model in order to reduce its complexity. A way to more closely follow the model without introducing unbearable experimental complexity must be addressed in the future.

5 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 concerned the experimental exploration of the complex relationships between factors such as the ability and integrity dimensions of trustworthiness, risk, sanctions and betrayals. From the gained experience in this work, we concluded that the conjunctive use of trust and norms – described in theoretic works as a promising interrelated governance mechanism – is not a trivial task. Several variables exist that must be accounted for when relating sanctions and trust.

Even though our model of agents' behavior is more realistic than the simple static and probabilistic models that are usually implemented to (individually) evaluate normative and trust models, it is limited by the fact that it is not empirical. Therefore, the best it can do is to support our understanding about what is the best conjugation of trust and sanctions, in conditions similar to the ones generated by our model. However, the results obtained in these conditions confirm the theoretical suspicion that the interrelated use of trust and sanctions is beneficial in selection processes where interacting partners are not necessarily engaged in ongoing relationships.

As future work, we intend to explore different other ways of conjugating trust and sanctions in the process of drafting of sanctions. We also intend to explore other interaction scenarios, where the behavior of agents (trusters and trustees) is driven by their intrinsic ability, integrity, and benevolence factors.

6 Acknowledgments

This research is supported by Fundação para a Ciência e a Tecnologia (FCT), under project PTDC/EIA-EIA/104420/2008.

7 References

- [BACHMANN, 2001] BACHMANN, R.: Trust, power and control in trans-organizational relations. *Organization Studies* **22**(2), 341–369 (2001)
- [COLQUITT *et al.*, 2007] 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** (4), 909-927 (2007)
- [CVETKOVICH *et al.*, 2002] CVETKOVICH, G., SIEGRIST, M., MURRAY, R., TRAGESSER, S.: New Information and Social Trust: Asymmetry and Perseverance of Attributions about Hazard Managers. *Risk Analysis* **22** (2), Blackwell Publishing, Inc, 359-367 (2002)
- [DAS and TENG, 1998] DAS, T.K., TENG, B.: Between trust and control: Developing confidence in partner



- [ELANGO VAN and SHAPIRO, 1998] ELANGO VAN, A.R., SHAPIRO, D.L.: Betrayal of trust in organizations. *The Academy of Management Review* **23**(3), 547–566 (1998)
- [FINKEL *et al.*, 2002] FINKEL, E. J., RUSBULT, C. E., KUMASHIRO, M., HANNON, P.: Dealing with betrayal in close relationships: Does commitment promote forgiveness?. *Journal of Personality and Social Psychology* **82** (6), 956–974 (2002)
- [FITNESS, 2001] FITNESS, J.: Betrayal, rejection, revenge, and forgiveness: An interpersonal script approach. In: M. Leary (ed.) *Interpersonal rejection*, Oxford University Press, 73–103 (2001)
- [HARDIN, 2001] HARDIN, R.: Conceptions and explanations of trust. In: Cook, K. S. (ed.) *Trust in society* (2), Russell Sage foundation series on trust, 3–39 (2001)
- [HARDIN, 2004] HARDIN, R.: Distrust: Manifestation and management. In: R. Hardin (Ed.), *Distrust*, Russell Sage Foundation, 3–33 (2004)
- [IRELAND and WEBB, 2007] 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)
- [LEWIS and WEIGERT, 1985] LEWIS, J. D., WEIGERT, A.: Trust as a Social Reality. *Social Forces*, **63**(4), Oxford University Press, 967–985 (1985)
- [LUHMANN, 1979] LUHMANN, N.: *Trust and Power*. John Wiley & Sons, New York (1979)
- [MACY and SKVORETZ, 1998] MACY, M. W., SKVORETZ, J.: The Evolution of Trust and Cooperation between Strangers: A Computational Model. *American Sociological Review* **63** (5), 638–660 (1998)
- [MAYER *et al.*, 1995] 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)
- [POORTINGA and PIDGEON, 2004] POORTINGA, W., PIDGEON, N. F.: Trust, the Asymmetry Principle, and the Role of Prior Beliefs. *Risk Analysis* **24** (6), 1475–1486 (2004)
- [SAKO, 1998] 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)
- [TAN and THOEN, 2000] TAN, Y.H., THOEN, W.: An Outline of a Trust Model for Electronic Commerce. *Applied Artificial Intelligence* **14**(8), 849–862 (2000)
- [URBANO *et al.*, 2009] 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)
- [URBANO *et al.*, 2011] 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)
- [URBANO *et al.*, 2012] URBANO, J., LOPES CARDOSO, H., ROCHA, A.P., OLIVEIRA, E.: Trust and Normative Control in Multi-agent Systems: An Empirical Study. In: *Highlights on Practical Applications of Agents and Multi-Agent Systems, Series: Advances in Intelligent and Soft Computing*, **156/2012**, 207–214 (2012)
- [WATHNE and HEIDE, 2000] WATHNE, K.H., HEIDE, J.B.: Opportunism in interfirm relationships: Forms, outcomes, and solutions. *The Journal of Marketing* **64**(4), 36–51 (2000)
- [WILLIAMSON, 1979] WILLIAMSON, O.E.: Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics* **22**, 233–261 (1979)
- [YAMAGISHI and YAMAGISHI, 1994] YAMAGISHI, T., YAMAGISHI, M.: Trust and commitment in the United States and Japan. *Journal of Motivation and Emotion* **18** (2), Springer, 129–166 (1994)

