The Knowledge Engineering Review, Vol. 29:1, 31–50. © Cambridge University Press, 2013
doi:10.1017/S0269888913000064
First published online 21 March 2013

On the integration of trust with negotiation, argumentation and semantics

PIERO BONATTI[1], EUGENIO OLIVEIRA[2], JORDI SABATER-MIR[3], CARLES SIERRA[3] and FRANCESCA TONI[4]

[1]*Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Universitá di Napoli Federico II, Via Claudio 21, I-80125 Napoli, Italy;*
e-mail: bonatti@na.infn.it;
[2]*Department of Computer Science (DEI), Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal;*
e-mail: eco@fe.up.pt;
[3]*Artificial Intelligence Research Institute (IIIA) of the Spanish Research Council (CSIC), Campus UAB, 08193 Cerdanyola, Catalonia, Spain;*
e-mail: jsabater@iiia.csic.es, sierra@iiia.csic.es;
[4]*Department of Computing, Imperial College London, South Kensington Campus, London SW7 2AZ, UK;*
e-mail: ft@imperial.ac.uk

Abstract

Agreement Technologies are needed for autonomous agents to come to mutually acceptable agreements, typically on behalf of humans. These technologies include trust computing, negotiation, argumentation and semantic alignment. In this paper, we identify a number of open questions regarding the integration of computational models and tools for trust computing with negotiation, argumentation and semantic alignment. We consider these questions in general and in the context of applications in open, distributed settings such as the grid and cloud computing.

1 Introduction

Agreement Technologies refer to computer systems in which autonomous agents negotiate with one another, typically on behalf of humans, in order to come to mutually acceptable agreements (see Ossowski, 2008a, 2008b). Agreement Technologies include trust computing, negotiation, argumentation and semantic alignment.

Argumentation is a powerful technique aiming at the evaluation of possible conclusions/claims by considering reasons for and against them. These reasons (arguments and counter-arguments) provide support for and against the conclusions, through logical reasoning. Although originally proposed within the realms of logic, philosophy and law, in the last decade argumentation has attracted wide interest in computing to understand and meet the challenges of a number of applications characterized by the lack of certain, consistent and complete information, and when numerical (e.g. statistical) information is not available or is only partially available (e.g. see Besnard & Hunter, 2008; Rahwan & Simari, 2009).

Negotiation allows agents to try and reach satisfactory agreements, fulfilling the needs and requirements of all agents involved. Argument-based negotiation differs from other approaches to negotiation in that it makes use of offers, counter-offers and refusals supported by arguments that explain the reasons behind them. In contrast with game-theoretic or heuristic negotiation (see Rahwan *et al.*, 2003), agents involved in argument-based negotiation may change their preferences, utilities, and even issues and goals on-the-fly.

Semantic alignment of ontologies aims at resolving the issues arising from distributed and independent knowledge creation and management. In an open framework where different agents

may use different ontologies to drive their behaviour and interact with one another, it is crucial to set up methods and mechanisms to find semantic relationships among syntactically unrelated concepts, in order to reach a shared mutual understanding of requests, offers, constraints, goals, etc. (e.g. see Noy, 2004; Choi *et al.*, 2006).

Trust measures are responsible for guaranteeing security on execution and help agents to determine with whom to interact and what terms and conditions to accept as a basis for the interaction (see Sierra & Sabater-Mir, 2005). Trust is usually understood as expectation on behaviour post-commitment (see Debenham & Sierra, 2006). Usually agents establish agreements that are later on executed, but this execution might not be as expected due to many reasons, such as execution or communication errors, defection or misunderstandings, just to put a few. Reputation is usually defined as a group's opinion on some aspect of an object (see Sierra & Debenham, 2009). For instance, the group may consist of buyers, the object may be a seller, and the aspect of interest may be the quality of products sold by the seller. Trust and reputation are closely related as trust, built along time, is the basis to express an individual opinion later aggregated into a reputation. Moreover, a reputation value may be used to build trust.

In this paper, we consider several open questions concerning the integration of trust (and reputation) computing techniques and other Agreement Technologies (negotiation, argumentation and semantic alignment), in the context of existing work. The open questions identify trends and challenges for future work in this area. In particular, we focus on the integration of trust and negotiation models (Section 2), the integration of trust and argumentation models (Section 3), the use of semantic alignment techniques for trust (Section 4), as well as current practical bounds (Section 5) and challenges (Section 6) for these Agreement Technologies in practical applications and in particular in open, distributed settings such as the grid and cloud computing. Finally, we conclude in Section 7.

2 Trust and negotiation

Trust/reputation and negotiation models can be integrated to benefit one another in (at least) three different ways:

- the outcome of negotiation can be a set of digital credentials, supporting the trustworthiness of potential partners in interaction and contributing to their reputation;
- trust/reputation models can help determine the right partners prior to or as a result of negotiation;
- trust/reputation models can help improve the outcome of negotiation.

We consider these three aspects of the integration in turn.

2.1 Trust negotiation

The integration of trust and negotiation models has already been profitably worked out in the area of computer security and privacy, in the form of *trust* (or *credential*) *negotiation* (e.g. see Winslett *et al.*, 2002; Staab *et al.*, 2004). Trust negotiation is based on the automated negotiation of digital credentials and other forms of evidence in order to implement attribute-based access control without placing the whole burden of information exchange on the shoulders of users. The following is a classical example of trust negotiation. Alice visits for the first time an online bookshop that sells an interesting book at a very cheap price. By interacting with the bookshop's server, Alice's personal agent finds out that she has to provide either her credit card number or a pair (*userId, password*) for a previously created account. Alice does not want to create a new account on the fly, so releasing her credit card is the only option. However, she is willing to give such information only to trusted online shops (e.g. belonging to the BBB[1]); therefore, her personal agent asks the bookshop to provide information on its membership to the BBB. The bookshop

[1] `http://www.bbb.org/`

belongs indeed to the BBB and is willing to disclose such credential to anyone. This satisfies Alice's policy, so her personal agent provides her credit card number and the transaction is successfully completed.

Some of the main advantages of trust negotiation are:

- privacy enhancement, because it is not required to disclose the full identity of the user each and every time; access control can be based on selected properties, such as age, nationality, etc.;
- digital credentials constitute a more reliable kind of information, as opposed to the unverified profiles associated to internet accounts;
- improved user experience, as the operations related to security and privacy become more transparent, and server properties (e.g. certifications) can be automatically checked, thereby making navigation safer.

Trust negotiation is a well-established area: many interesting approaches and results already exist (e.g. see Bonatti *et al.*, 2009), so the real question is what still needs to be done. One important issue concerns the design of privacy-enhancing negotiation strategies that effectively reduce the amount of disclosed information. In this respect, strong guarantees are hard to achieve given the lack of trusted third parties on the Web, and given the lack of mutual knowledge among the negotiating peers. Market design mechanisms seem a promising direction, as a way to establish a system of incentives and disincentives that tend to minimize the amount of user information requested by servers for access control (see Bonatti *et al.*, 2011). Another interesting open issue is how policies themselves can be negotiated, possibly relaxing some requirements in a controlled way, in order to increase the number of successful negotiations and get compensation for unsatisfied security/privacy preferences.

2.2 Trust/reputation models to select partners in negotiation

One of the standard uses of trust and reputation models is to allow an agent to choose the right partner for an interaction. In this section, we focus on situations where this interaction is a negotiation, for example, for (one or more) services in grid computing or in cloud computing settings.

One of the main aspects that a negotiator, being it human or software, needs to assess in a negotiation process is the true capabilities of any potential partners in joint actions (see Sierra & Debenham, 2007). The capabilities are of two fundamental types: functional and non-functional. When joint actions amount to the execution of software services, functional capabilities refer to the input/output relationship of software execution, that is *what* the software can actually compute. Non-functional capabilities refer to *how* the computing is actually executed, for example how long it takes, what percentage of errors are produced, how many resources are required, etc.

If we understand trust as expected behaviour after commitment (e.g. as in Debenham & Sierra, 2006), it is clear that trust measures can be used to assess both types of capabilities: functional, because the input/output relationship is a commitment (specification) on behaviour, and non-functional as well, because negotiations involve quality of service agreements that are commitments that can later on be observed. For instance, if a service commits to compute a $n \times n$ matrix inversion in $n^2 \times 200$ ms, the client can check whether that is the case and learn a model of service behaviour that will help the client in assessing the trust on future service commitments. This assessment will determine whether to negotiate with the service in the future and what service-level agreement (SLA) to agree upon.

Using reputation models to evaluate a partner as a negotiator is not very different from using reputation models to evaluate any other characteristic of that partner. In this sense, current reputation models (for example, models like those of Mui *et al.*, 2002; Regan & Cohen, 2005; Huynh *et al.*, 2006) are suitable for this task. The integration with the negotiation model implies, among other things, identifying the set of features that typify a negotiator and establishing a classification of types of negotiators. Several issues are still unresolved/ongoing in this setting. We identify a number of these below.

2.2.1 Commitments and trust

For trust to be assessed in the context of commitments, a suitable representation and ways for determining fulfilment of commitments are needed. These may need to take into account 'mitigating circumstances' (in the sense of Miles *et al.*, 2009b) that may have determined or contributed to the non-fulfilment of commitments and should not contribute negatively towards the assessment of trustworthiness of an agent. Also, events that have positively impacted on fulfilment of commitments but are outside the control of the target agent should not contribute positively towards the trustworthiness of agents. We will discuss, in Section 3.1, how argumentation can support some of this reasoning.

2.2.2 Context and trust

Trustworthiness may need to refer to specific aspects of the behaviour of agents (or of the services they provide). For example, an agent may trust a service provider as far as the functional properties of the services it provides are concerned, but not about some non-functional property (e.g. timeliness of the delivery). Then, the selection of an agent to interact with becomes a multi-attribute decision-making problem. Information about the trustworthiness, or lack thereof, of an agent concerning specific aspects may determine the negotiation strategy of the other agents (see also Section 2.3). For example, an agent may insist on a strongly binding SLA for an aspect it does not consider the interlocutor agent trustworthy about.

Also contextual attributes that characterize the deal under negotiation should play an important role on both how to build up trust and reputation images and how to use them for negotiation. A real useful computational trust and reputation model, more than providing one or two indicators on whether to trust a potential partner, has also to provide the contexts for which those indications are given. For example, a potential partner may be good at supplying some good G in general. However, if this partner is hurried to deliver G in a very short time, it may fail to deliver G appropriately. Therefore, classification and clustering inspired algorithms can be used to attach trust measures to agents and select them according to the current context conditions so that they can be accurately used during specific contract negotiation processes. By using context-based features (in the form attribute-value), metrics based on the frequency of these features can be defined, such as the gain criteria of Urbano *et al.* (2010c). Then, suing any such metric, clusters of agents that behave similarly in given situations can be formed.

Moreover, the respective stereotypes, characterizing those created clusters, can be derived giving a much finer-grained perception on agents' behaviour, as well as fair trustworthiness evaluation. Decision making on partner selection becomes, through this process, more tuned with current context. For example, agents belonging to a specific cluster may prove to violate contracts whenever delivery time is low and, thus, despite a generic acceptable trust measure, they should not be selected in such a context (as suggested in Urbano *et al.*, 2010c). Also information gain-based methods are being successfully tried out to improve accuracy of tendency extraction from agent behaviour, contributing to a more reliable online, situation-aware trust model (see Urbano *et al.*, 2010c).

2.2.3 Bundles and trust

If a combination of items, for example, services, is negotiated with several agents, then there is a key issue that needs to be addressed and where trust plays a role: *incompatibility between services*. Incompatibility may derive from semantic mismatches between the specification of the services or due to differences in non-functional requirements (e.g. time, quality, optimality). On the latter aspect, trust may determine the overall level of compatibility of the final solution, by computing a level of expectation over the non-functional behaviour of the solution by using previous experience with each service and by looking into the concrete orchestration and choreography of the services.

2.2.4 Groups and trust

There are two aspects that are relevant from a trust perspective when agents team up to solve a particular problem. First, the *composition* of a team: agents that may be trustworthy in isolation

may turn out to be untrustworthy when teaming up with other agents, or, conversely, agents that are not to be trusted in isolation may be fine when teamed up with others. Thus, the composition of a team turns up to be a complex combinatorial problem. Second, the trust *aggregation*: as a consequence of the first issue, the trust on a group as a whole is not functional, that is, it is not a simple aggregation of the trust on the individuals. Information about the compatibility between members is required to perform this calculation. Again, a combinatorial problem appears.

2.2.5 Negotiation skills and trust

If an interaction involves a negotiation process, choosing the right partner should imply not only the evaluation of the partner from the point of view of the service it can offer but also the expected characteristics of the negotiation that will take place before agreeing upon the service. Indeed, it is worthless for an agent to select a partner that hypothetically can offer the best service if it is known *a priori* that the interaction will never happen because the negotiation process will fail. This failure in the negotiation stage can be for different reasons and some of them can be foreseen by using trust and reputation models. One possible reason for failure is the lack of suitable negotiation skills. For example, it may be the case that an agent's negotiation mechanism is too simple to obtain a good deal when the agent interacts with partners with greater negotiation skills. Knowing the reputation of the possible partners as negotiators can help an agent to balance (i) the quality of the service they potentially can offer with and (ii) the actual possibilities for the agent to obtain that quality after a negotiation process.

2.2.6 Trust, reputation, institutions and norms

Whenever there is a need for different entities to negotiate possible joint activities, agents seeking information about potential partners may trigger other agents into forwarding noisy as well as false information, depending on the degree of competitiveness involved in that particular negotiation. This implies that we should be very cautious about how, when, and what for information on trust and reputation is produced and used.

When negotiation between different entities implies some kind of competition, available information about different negotiation players in the scenario needs to be clearly backed by the system/institution through which the negotiation process takes place. A clear separation has to be enforced between what is, on one hand, a more reliable situation-based (*context sensitive*) trust and, on the other hand, socially generated information (reputation). The former is typically local, sporadic, associated with an agent-specific functionality, activity or even role; also, it is mostly dynamic, in the sense that it decays over time unless new evidence arrives (see Melaye & Demazeau, 2005), it is asymmetric, in the sense that it grows slowly and decays fast, and it embeds the 'maturity property', meaning that trust changes may be slower after a higher degree of trustworthiness has been reached (see Urbano *et al.*, 2009). Reputation, on the other hand, relies less on the agents' own experience, is broader in scope, sometimes fuzzy, and can only be acceptable if and only if it obeys some explicit, institutional, social norms governing the propagation of trust-based information in a network. Such norms should prevent unfair generalization and abuse of the transitivity property (as discussed in Tavakolifard *et al.*, 2009), only allowing well-formed reputation information to be taken into account.

Agents interoperability, negotiation, cooperation and joint work become more credible if regulated by institutional norms. Norms, prescribing obligations, sanctions and, possibly, rewards, are enforced to entities (agents) by some kind of accepted institutional power (electronic institution). Usually the outcome of a negotiation is a (formal or informal) contract, whose clauses may trigger, in the normative environment, appropriate-specific norms for that particular context. According to a model inspired by contract law theory (see Craswell, 2000; Kaplow, 2000), prescribing the use of norms and default rules for different contexts, we intend to benefit from trust and reputation measures to select appropriate norms that have proven to be more adequate to situations where specific types of (more or less reputable) agents are present. This will lead the normative environment to adapt to the current situation, by applying norms appropriate to each

specific context, trying to be fair according to what exactly has been perceived before, about the contracting agents (see Cardoso & Oliveira, 2008a, 2008b).

Moreover, besides being a kind of mediator on sanctioning violation of mutual agreed contractual clauses, a normative environment can also enforce punishments (through fines, for example) in some specific situations. This is particularly important in application domains (like B2B) in which either bi (or multi) lateral relationships easily produce social effects, leading to the need for a broader recognition of those effects. Thus, peer-to-peer relationships may not affect just those agents directly involved, but also either all or a subset of the society they belong to. That is why, besides direct material sanctions, leading to a direct compensation of the counterparty, there is often the need for the society to enforce fines in order to discourage future violations. Deterrence transforms an institution from a mediator into an enforcing mechanism whenever (and only when) this is considered good for the society (see Cardoso & Oliveira, 2008a, 2008b).

Also, whenever an agent is not completely sure (or is even suspicious) about the other partner's future behaviour (in the case of a newcomer, for example), it may be reassuring to know that there is a normative environment in charge of monitoring the flow of events and being responsible for enforcing the pre-existing applicable norms. Moreover, the concept of a norm, although mainly related with prescribing sanctions, can/should be enlarged to encompass incentives (rather than just punishments or sanctions) to those who behave well, thus favouring correct behaviours. In this way, a normative environment can be responsible for encouraging and stimulating, through social behaviour monitoring and norm enforcing, fair social behaviour.

Incorrect broadcasting, through computational trust and reputation mechanisms, of one agent's perception on other agents' trustworthiness that could lead to incorrect bad reputation propagation, has to be monitored and sanctioned by the normative environment. These procedures will entangle the computational trust and reputation mechanism with both the partners selection process and the monitoring capabilities of a normative environment, leading to more fairness in the application of norms.

2.3 Trust models to improve the outcome of negotiation

In the previous section, we have seen how trust models can help choosing a partner. Once the negotiation process has started, the trust and reputation model can play a new role, of conducting or at least influencing how the negotiation process takes place. This use of the trust and reputation model, not as a mere partner selection mechanism but as a piece of the decision-making module, implies a tighter integration of the model with the other elements of the agent architecture, something that in current models has been somewhat neglected. Moreover, the integration has to be also made at the organizational level, involving the trust and reputation model in the decisions about communication and social-related actions.

In the remainder of this section, we explore these two uses of trust and reputation models: to direct negotiation and to improve contracts resulting from negotiation.

2.3.1 Trust/reputation to direct negotiation

Once the negotiation process starts, all the knowledge agents have about the negotiation strategies used by their partners can help to improve the outcomes of the negotiation. Again, the reputation of a partner as negotiator (not only how good it is at negotiating but also the kind of negotiation mechanism it uses) can be useful to direct the negotiation process.

By using context-based agents' trustworthiness, an agent can either pre-select for negotiation the right potential partners or, during the negotiation process itself, help on selecting the best proposals for each particular situation (see Urbano *et al.*, 2010a). Once again, it is worthwhile to state that to jointly work safely, agents need to go beyond generic indications given by trust and reputation measures, and have to analyze other agents' trustworthiness for each particular relevant context.

For example, if our partner has the reputation of trying to force, always a quick negotiation because it has a very good algorithm for a specific kind of negotiations, we can decide to slow

down the rhythm of the negotiation hoping for possible mistakes in our favour. Or if we cannot force the negotiation conditions, we can simply use the reputation information to choose the right negotiation algorithm (in this case a fast and probably simple algorithm) that allows us to be competitive. Of course, the context plays a very important role in this use of reputation for negotiation. The selected negotiation strategy usually will not only depend on the reputation of the partner as a negotiator but also on the context surrounding that particular negotiation. In our previous example, we can probably slow down the rythm of the negotiation only if we have a stronger position than our partner.

2.3.2 Context-based contract negotiation

Normative environments (see Cardoso & Oliveira, 2005, 2008b) may impose to consortia resulting from negotiated agreements *contracts* to be of different types according to partners' degree of mutual trustworthiness. Partners with lower credibility may get tougher electronic contracts including more strict clauses and harder sanctions in case of commitment failure.

When trust is used for partner selection (see Section 2.2), trust values can be considered as hard constraints (e.g. a prerequisite for B2B partner selection processes) prior to negotiation. However, simultaneously, trust values can be seen as a kind of soft constraint influencing and driving a (possibly argumentation-based) negotiation process whose outcome is the final decision on whether to work jointly or not. This is often the case in real B2B domain negotiations. In this setting, a progressively more detailed information exchange, starting with asking information about the team and past partners, may end by asking relevant, and sometimes more private, information (e.g. financial) together with specific guarantees about future behaviour. This means that 'proposing/counter-proposing' goes along with 'asking for/answering to' certain kinds of guarantees closely related with the importance of what is being negotiated at that particular moment and the involved calculated risk.

Pre-existing information (e.g. previous contract outcomes in a B2B scenario) can be used through conceptual *clustering* (non-supervised learning) algorithms to derive stereotypes, namely sets of the most discriminating features characterizing the elements of each cluster, and better map agents' trustworthiness according to the different contexts (i.e. new contracts under negotiation, see Urbano *et al*., 2010b, 2011). Moreover, associating meaningful stereotypes to clusters of previously known agents (representing entities, for example negotiation partners) may be of great help when dealing with *newcomers* trying to temporarily include them in known clusters, and deducing some kind of expectation on their future behaviour without being either too unfair or too naive. When more interactions take place and more information becomes available, these more recent known entities can either confirm or deny their previous classification.

Computational trust and reputation models have already been included in electronic institution platforms, mainly for regulating the virtual enterprises life cycle (see Cardoso & Oliveira, 2005; Urbano *et al*., 2009). However, we envisage, for the future, that these modules will be at the heart of the agents' decision making whenever they decide to communicate in the framework of a social network (see Lacomme *et al*., 2009). This will permit us to evaluate which kind of relationships have been established in the network, regarding stability, openness, specific nodes criticality and centrality, sub-networks aggregation or disaggregation forces, etc. Playing with different computational trust and reputation models may lead to very different (social) networks, with different characteristics. Conclusions will thus be drawn about what features should be included in the trust and reputation models that may lead to either more socially aware or risk tolerant social networks.

3 Trust and argumentation

Argumentation can be used to *justify* an observed behaviour, and therefore modify the impact on the model of expected behaviour that is at the base of trust models (see Sierra & Debenham, 2009). For example, a partner in an interaction could provide an evaluator with a justification of the form 'I could not deliver the package yesterday because the driver had a flu'. The recipient of the

information may decide to believe that and thus not alter (decrease) the trust on the interaction partner (the courier service company in our example).

There are (at least) three forms of integration, beneficial to either or both models:

- a trust model can be used to determine the reliability of an argument;
- an argumentation process can be used to improve the performance of a trust model;
- argumentation can be used in support of trust negotiation (see Section 2.1).

We consider these three aspects of the integration in turn.

3.1 Trust for argumentation

3.1.1 Provenance and argumentation

During an argumentation process, there is an exchange of arguments that are built using different pieces of knowledge. Whether an argument is finally accepted or not depends, among other things, on the structure of the argument and the 'truth' behind the knowledge used to build the argument. More than often, the 'truth' of the knowledge that is used to build the arguments may depend on the source of that knowledge. The stream of information received from an informant together with its reputation are the only elements an agent (or human user) can use to decide whether an argument, built from that information, can be accepted or not.

The use of trust models for this purpose is straightforward. Currently, one of the main uses of trust models in multi-agent systems is the evaluation of a piece of information regarding the source (who is the origin) of that information. Here the current work on representation of provenance can be of great importance (see Miles *et al.*, 2009a; Moreau *et al.*, 2011).

Extending this kind of evaluation to each one of the elements of a (structurally correct) argument, we could evaluate the truthfulness of that argument. It would be interesting to explore definitions of acceptability of arguments using a combination of standard dialectical acceptability and trustworthiness/reputation of the source of the argument.

3.1.2 Social relations and argumentation

The importance of social relations in the acceptance of arguments has been long studied by psychologists (e.g. see Karlins & Abelson, 1970). For instance, arguments produced by someone holding the authority usually produce a bigger change in the beliefs of the hearer. Also, belief changes persist longer if the argument is generated by a peer. Exploring the social dimension seems an interesting line of work that has not been studied in depth yet.

3.2 Argumentation for trust

A different but complementary approach is to use argumentation to improve the performance of the trust (or reputation) mechanisms. This can happen when trust is computed by an agent in isolation (relying on past interactions with a target) or when agents can exchange and share information about the trustworthiness of possible targets (or one another).

Computing trust is a problem of reasoning under uncertainty, requiring the prediction and anticipation by an agent (the evaluator) of the future behaviour of another agent (the target). Despite the acknowledged ability of argumentation to support reasoning under uncertainty (e.g. see Krause *et al.*, 1995), only Prade and Subrahmanian (2007), Dondio and Barrett (2007) and Parsons *et al.* (2010) have considered the use of arguments for computing trust in a local trust rating setting. Dondio and Barrett (2007) propose a set of trust schemes, in the spirit of Walton's argument schemes, and assume a dialectical process between the evaluator and the target whereby the evaluator poses critical questions against arguments by the target concerning its trustworthiness. Prade and Subrahmanian (2007) proposes an argumentation-based approach for trust evaluation that is bipolar (separating arguments for trust and for distrust) and qualitative (as arguments can support various degrees of trust/distrust). Parsons *et al.* (2010) define an argumentation logic where arguments support measures of trust, for example, qualitative measures such as 'very reliable' or 'somewhat unreliable'.

There are several non-argumentation-based methods to model the trust of the evaluator in the target. Sabater and Sierra (2005) classify approaches to trust as either 'cognitive', based on underlying beliefs, or 'game-theoretical', where trust values correspond to subjective probabilities and can be modelled by uncertainty values, Bayesian probabilities, fuzzy sets, or Dempster–Shafer belief functions. The latter approach is predominant nowadays for trust computing. However, Castelfranchi and Falcone (2000) argue against a purely game-theoretic approach to trust and in favour of a cognitive approach based upon a mental model of the evaluator, including goals and beliefs. Moreover, some works (e.g. see Staab & Engel, 2008) advocate the need for and benefits of hybrid trust models, combining both the cognitive and game-theoretical approach.

Recent work by Matt *et al*. (2010) proposes a hybrid approach for constructing Dempster–Shafer belief functions modeling the trust of a given agent (the evaluator) in another (the target) by combining statistical information concerning the past behaviour of the target and arguments concerning the target's expected behaviour. These arguments are built from current and past contracts between evaluator and target, and are integrated with statistical information proportionally to their validity. The method extends a standard method for trust (namely that by Yu & Singh, 2002) that relies upon statistical information only. The two methods have identical predictive performance when the evaluator is highly 'cautious', but the hybrid method gives a significant increase when the evaluator is not or is only moderately 'cautious', where the level of 'cautiousness' of the evaluator is understood as a measure of how risk-averse the evaluator is (the higher this level is the more risk-averse the evaluator is). Moreover, with the hybrid method, target agents are more motivated to honour contracts than when trust is computed on a purely statistical basis.

3.2.1 Argumentation profiles and trust

When agents engage in argumentation dialogues they show different profiles. For instance, an agent may be more persuasive than others in the sense that the final opinions of the others vary more significantly, or it may predetermine the result of the argumentation process as the final joint opinion is closer to its initial position than to the others'. Given that trust is usually related to expected behaviour, the composition of a team or a committee where the members have particular profiles is key to assessing the trustworthiness of the team in reaching decisions. This aspect of argumentation has not been studied in detail. Some initial ideas in the area of reputation can be found in the work by Sierra and Debenham (2009) and in the reputation model of the LiquidPub[2] project.

3.2.2 Trust information exchange

Agents can exchange trust/reputation information in the form of arguments, for example, as in the work by Lenzini *et al*. (2008). One of the key aspects in argumentation (see Sierra & Debenham, 2008) is *when* and *whom* to disclose *what* type of information. The behaviour of agents, services and humans is just one type of such information that can be disclosed. The information that 'My butcher gave me shocking chops yesterday!' may persuade others to buy somewhere else. This information may be incorporated in trust models of others depending on their interpretation of the information and their trust in turn on the agent providing the information (see Sierra & Debenham, 2009).

3.2.3 Communicating justifications

Instead of just communicating an isolated trust/reputation value, an agent can communicate the value plus a justification of that value. After that, an argumentation-based dialogue can help to decide if that trust/reputation value is worth to be added to the agent's mind as new knowledge (see Pinyol & Sabater-Mir, 2009). Given that, how the reputation model calculates the reputation value and therefore how this justification is built becomes very relevant. If we want that agents and humans share common virtual spaces and therefore interact among them, using a cognitive

[2] www.liquidpub.org

approach in the design of reputation models is the best approach. In these kind of environments, we need agents that can build justifications like humans so other humans can easily understand those justifications. The agents have to be able to understand also arguments given by humans. To achieve that, from a computational point of view we need several elements:

- *A cognitive theory on reputation*: This is the starting point and for this, we can rely on social sciences that have studied reputation from a cognitive perspective for many years. A good example of cognitive reputation theory is that from Conte and Paolucci (2002).
- *A computational model of that theory*: At the end, we want the reputation model to be into an agent's mind so the initial reputation theory has to be formalized and implemented.
- *A language that allows to express the concepts related to reputation as logical predicates*: The agents have to be able to analyze the justification and reason about it.
- *A framework that integrates everything in a cognitive architecture*: Although simple reputation models can be considered in isolation and therefore the justification can be built only from internal elements of the model, once one moves to more complex reputation models the connection with other modules of the agent needs to be taken into account. Trying to justify properly a reputation value without taking into account the plans or the goals of the agent is not feasible. A proper justification will require to follow the justification thread outside the reputation model. Because we want a cognitive justification, having a cognitive agent architecture that allows to follow a justification thread outside the reputation model is the best approach.
- *A reputation ontology and a language based on that ontology to talk about reputation*: At this point, the agent is able to build a justification taking into account not only the elements strictly attached to the reputation model but also to other relevant modules of the architecture. Now it has to be able to communicate that justification and therefore a language and a common ontology are required.
- *An argumentation framework and a dialogue protocol to allow agents to seek information*: We have the justification, we have a language that is understood by the agent's partners and can be used to communicate the justification. The last thing needed is an argumentation framework and a dialogue protocol that allow the agents to argue about that reputation value.

The final result of all this process is that the agents should be able to exchange reputation values, justify why they think those values are right and discuss with others about them, changing their minds if the arguments are convincing enough. Although complex, all this process can considerably improve the usefulness of reputation values, giving them an extra level of reliability.

3.2.4 Ignorance and trust

A negotiation system, and an electronic institution 'supervising' the negotiation process, should know and be clear about the contexts in which it is possible for entities to provide faithful information on trust. Moreover, an electronic institution should be able to prevent trust measures from being formed and shared based on misleading grounds. This implies that there are situations in which trusting/not trusting are not the only options. Indeed, some kind of measure of 'ignorance' should be computed (as it is the case with the use by Dempster–Shafer models of the plausibility concept). An argumentation-based process can then be followed for the sake of decreasing the current ignorance if this is higher than a threshold, in order to reach a situation in which agents cannot, at that time, progress much more. Agents have thus to decide if either their trust on an entity (agent) is reliable or not. Normative environments (such as those by Cardoso & Oliveira, 2005) will play here a role in a way similar to a kind of anti-spam system by preventing ill-formed information to be taken into consideration.

3.3 Argumentation for trust negotiation

Argumentation has not (yet) been applied in the context of trust negotiation. In most trust negotiation systems, access control and credential disclosure decisions are actually based on proofs

that constitute arguments in favour of permitting the incoming request. Negative decisions usually result implicitly from the failure of all policy rules (i.e. there are no counter-arguments).

In the broader literature on computer security, one can find security models supporting positive and negative authorization inheritance with exceptions, which more closely resemble an argumentation system: the proof of a specific (positive or negative) authorization (e.g. 'staff can access confidential files') may defeat a recommendation for a more general case (such as 'users cannot access confidential files'; see also the survey by Bonatti & Samarati, 2003). Conflict handling is specified via explicit rules, see for example Jajodia *et al.* (2001).

3.3.1 Extending trust negotiation with negative authorizations and conflict resolution

In principle, in trust negotiation there is space for such nonmonotonic mechanisms because (i) authorization overriding provides an appealing way of formulating policies incrementally, through iterative refinements, and (ii) different contexts require different forms of conflict resolution, therefore a flexible, rule-based approach is attractive.

Such extensions, however, raise at least three open questions. First, it is not yet known how negation and conflict resolution would affect the computational complexity of automated trust negotiation.

Second, a similar question applies to negotiation *strategies*. It is not only a matter of complexity: in the literature, strategies are usually required to satisfy an *interoperability* property (strategies should not cause unnecessary negotiation failures) that currently has been proved only in the absence of overriding and conflict handling (by Yu *et al.*, 2001, 2003; Baselice *et al.*, 2007); how would the latter interfere with strategy interoperability?

Third, the impact of the new features on usability—for example, due to the subtleties of nonmonotonic reasoning—should be systematically investigated, as we know that lay users find it difficult to use even extremely simple and streamlined policy languages, see, for example, the user studies reported by Sadeh *et al.* (2009). Intelligent authoring tools (like the policy learning approach of Sadeh *et al.*, 2009) may help to address such problems, however, we are only at the beginning of this line of research, and again the more powerful the policy language, the more difficult the development of effective authoring tools.

3.3.2 The potential of argumentation: unreliable knowledge and policy negotiation

So far, we have only touched upon integrations of argumentation and trust negotiation that—to a certain extent—are limited to extending trust negotiation with mechanisms that are already supported in logic-based policy frameworks without negotiation. However, such rudimentary forms of argumentation do not address a number of interesting issues that naturally arise in modern Web scenarios, where general argumentation techniques may bring significant benefits and show their full potential.

First, consider that the quality of knowledge may be highly dishomogeneous across different sources. Some works report systematic knowledge representation errors that trigger inferences that are logically correct (w.r.t. the encoded knowledge) but do not reflect the true state of affairs (see Hogan *et al.*, 2009). So a natural question is: can argumentation notions such as attacks and counterattacks be enriched with provenance, reliability and authoritativeness meta-information in such a way that argumentation can be used to address the problem of policy enforcement in the presence of unreliable and noisy knowledge? By comparing argument reliability and authoritativeness, it may be possible to improve the quality of (possibly distributed) policy-related decisions. Furthermore, similar ideas can be exploited to improve efficiency by restricting reasoning to a reliable subset of the relevant sources available on the Web, as in the work by Hogan *et al.* (2009).

Another topic—of great potential commercial interest—concerns extending automated trust negotiations with a bargain-like mechanism that may increase the number of successful transactions by progressively relaxing agent policies, for example, by trading service improvements for more information about users. In this perspective, policy rules are treated like soft constraints that may be relaxed if this brings sufficient benefits. Again, traditional trust negotiation mechanisms

are too rigid for this kind of *policy negotiation*, and the role of argumentation techniques in evaluating the pros and cons of relaxing a policy is still to be explored.

4 Trust and semantics

Trust information is 'linguistic'—thus semantics is essential. It is linguistic in the sense that behaviour is 'described' by words. For instance, if we need to describe the input/output relationship of a service, the input and output descriptions may not be equally understood by everybody. Semantic alignment permits to solve potential meaning mismatches. Also, when extracting information from past negotiation experiences, semantic similarity may help in extracting the most of it. For instance, if the promised lamb chops appeared to be excellent then the kid leg I am offered might be similarly good. Semantics and ontologies are fundamental in human trust modelling and should be equally important in automated trust models.

4.1 Semantics for trust

Semantics are fundamental for trust and reputation models as they are deployed in open environments where agents designed and implemented by independent developers have to interact and use trust and reputation mechanisms to increase the robustness of the overall system. In this setting, we cannot guarantee that all agents will use the same trust and reputation mechanism. In fact, the current situation is that there is a big diversity of models, each approaching the problem from a different perspective. These models frequently even differ on basic concepts such as what trust is and what reputation is. On the other side, trust (and especially reputation) models often require that there is communication among agents and exchange and sharing of information (as we discussed in Section 3.2.2). Given these considerations, it is clear that some kind of semantic alignment is necessary to allow the agents to exchange information about trust and reputation.

A straightforward approach to the problem is the use of a common ontology that can work as a bridge between the different models. Although some attempts to define such an ontology can be found in the literature (see Casare & Sichman, 2005; Pinyol *et al.*, 2007), it seems quite difficult to finally obtain an ontology that can be accepted by all the scientific community and that can give support to the different levels of complexity required by the different models. For example, there are models that use labels to represent reputation values while others are using fuzzy numbers or probability distributions. A possible alternative (currently under consideration by Koster *et al.*, 2010) is to use dynamic semantic alignment. In that case, the alignment is performed for each specific interaction. This would make possible to tailor an alignment that exploits much better the characteristics of the models being aligned. Whatever the approach, what is clear is that some kind of alignment is necessary to allow heterogeneous models in the same environment.

Trust arises from the analysis of a wide range of evidence of various strength, incrementally gathered during a sequence of interactions (e.g. transactions) in direct as well as indirect ways (like reputation systems do, for example). Semantic techniques address relevant problems such as recognizing problematic agents and combining heterogeneous information. Remarkably, in some trust negotiation systems like Protune (see Bonatti & Olmedilla, 2005; Bonatti *et al.*, 2010) agents exchange their requirements by disclosing selected parts of their policies in the form of logic programming rules. Such *shared knowledge* is a semantic technique in itself: it allows agents to interoperate through a knowledge-based formulation of their policies grounded on concrete data items such as X.501 digital credentials. All that agents need to understand *a priori* in Protune is the meaning of three predicates, because all the other application-dependent and domain-dependent abstractions adopted in Protune policies can be defined in terms of those three predicates, in a machine-understandable way. While this is an interesting proof of concept, showing that an ontology infrastructure is not necessarily a prerequisite for interoperability in a trust negotiation system, we remark that in the domains where such a semantic infrastructure is available, semantic

policies can benefit from the use of the available ontologies. Hybrid formalisms (integrating rules and description logics) are going to play an essential role in this area; researchers on policy languages should evaluate the many existing approaches w.r.t. complexity and expressiveness, taking in due account the specific requirements of policy languages.

Finally, when negotiation is integrated with information exchange to ascertain trustworthiness (see Section 2.3), a shared ontology of concepts and properties in the domain is required. When this is not available, semantic alignment techniques are needed.

4.2 Trust for semantics

A much less explored area of research is the use of trust techniques for semantic purposes. In particular, there are two research venues that require further attention: development of *semantic services*, for example for making alignments, and trust on *semantic resources*, for example trust on an ontology.

The development of semantic services seems unavoidable in an open world where the ontologies are being developed in large quantities. When orchestrating services to form complex solutions, semantic services will be required to make the glue among them (e.g. align ontologies of different services) as well as to facilitate the choreography between services that communicate using different messaging languages.

When querying ontologies, the answers' quality will depend among other issues on the provenance of the ontology. For instance, answers to queries on astronomy will probably be more trustworthy if coming from an ontology developed by NASA. Thus, the concept of *trust on an ontology*, or on combinations of ontologies, takes shape as a new concept that requires deeper analysis.

Of course, knowledge quality is not the only relevant parameter in this context; the methods and the algorithms that implement a semantic-based service are equally important, as well as the related issue of how much a particular service or agent can be trusted. A natural approach consists in evaluating a trust measure by observing the behaviour of a service or agent across a sufficently long sequence of transactions. However, for this purpose, it should be possible to tell whether two or more transactions involve essentially the same agent; this may be difficult when agents can clone or masquerade themselves. This problem has been partially addressed by the security community in several ways. For example, mobile code is usually signed by the producer and privileges are assigned to the code according to its signature (see, e.g. the latest Java security model of Gong & Ellison, 2003). In this case, trust is actually placed in the producer; however, the signature can be used to identify the code, as well.

Another possible approach to recognizing an agent across multiple transactions is *certificate fingerprinting* (see Lee & Winslett, 2006). Roughly speaking, the idea is that even if an agent (or a user) is not authenticated and transactions are guaranteed only by means of anonymous certificates, it is frequently possible to tell that a set of transactions has been performed by the same agent because some specific certificates have been repeatedly used in those transactions. Certificate fingerprinting actually identifies the agent's *principal* (i.e. the entity on whose behalf the agent is acting, that is the real owner of the certificates). In order to focus on agent recognition, certificates should be assigned directly to agents.

Whether these methodologies can be effectively applied for agent trust evaluation, and whether semantic techniques can improve their precision, are two interesting open questions that deserve further investigations.

5 Towards real-world scenarios

If Agreement Technologies, including negotiation, argumentation, trust and semantics techniques, are to be used in real-world scenarios, practical bounds to these techniques need to be removed. To do that there are two needs that have to be satisfied: testbeds to compare different approaches and efficient models that can scale.

5.1 Testbeds for trust

There are not many testbeds developed for trust in combination with negotiation, semantics or argumentation. We mention two of them that offer interesting opportunities.

5.1.1 Agent reputation and trust testbed

The agent reputation and trust (ART) testbed[3] (see Fullam *et al.*, 2005) is a testbed for computational reputation models designed to serve two roles: (1) as a competition forum in which researchers can compare their technologies regarding reputation models against objective metrics, and (2) as a suite of tools with exible parameters, allowing researchers to perform customizable, easily repeatable experiments. The testbed presents a scenario with simulated clients that request appraisals for paintings from different artistic eras. The participating agents have to provide these appraisals to the clients, however, they are not experts in all the areas so if an appraiser does not have the expertise to complete the appraisal, it can request opinions (paying accordingly for them) from other appraiser agents. The dilemma the agent has to confront is either to help other agents appraising as best as possible the requested paintings (and therefore wining a good reputation among appraisers as informer) or cheat them so the clients who ordered the appraisal to the requester agent are unhappy. The reputation model is the tool that the appraiser uses to take decisions at this level. Unfortunately, the ART testbed, though still available, is no longer maintained.

5.1.2 DipGame Testbed

The DipGame testbed[4] is based on the classical Diplomacy board game. The game is played by seven players incarnating political powers at the beginning of the 20th century with the goal of conquering Europe. It has the interesting property that there are no chance moves and that the main ability to win is to be a good negotiator. In addition, negotiations are private but decisions are public, therefore it is easy to verify whether secret pacts are kept or not, and therefore a model of trust on the other players can be built alongside a game, as there are two negotiation rounds per year (the game progresses along time) and games usually last for quite a number of years. The language used in the negotiation interchanges can be set to be as complex as desired (i.e. up to natural language) and therefore can potentially be used to also test argumentation models.

Powers can be incarnated by either humans or by softbots making the testbed ideal to compare negotiation strategies among themselves and with humans. The testbed offers facilities to build your own softbot in Java. See Fabregues and Sierra (2009) for details.

5.2 Scalability

The most urgent practical problem to address in trust and its relation with negotiation or semantics is scalability (e.g. how to reason over different ontologies and how to negotiate with different agents). In networks of thousands or millions of interoperating services/agents, a solution to scale up agreements is needed. Social techniques might be required to focus the search, around known and trusted neighbours, and semantics might be required to scale up the search for similar interests. Also, organizations might help by following a divide and conquer approach, clustering agents according to criteria like similarity in the preferences or in semantics.

By contrast, scalability is not a major problem with access control policies. Indeed, currently, real-world policies are not very large, partly because trust negotiation systems are not yet widely deployed; therefore, the application of articulated policies would place an excessive burden on users. Experimental result such as those reported by Bonatti *et al.* (2010) show that there are no immediate issues related to performance when using these policies. There are, however, serious difficulties related to usability. Recent studies show that users write very poor policies, even if the

[3] `http://megatron.iiia.csic.es/art-testbed/index.html`

[4] `www.dipgame.org`

policy language is extremely simple (see Sadeh *et al.*, 2009). An improper formulation of the user policy undermines the effectiveness of trust negotiation, and may cause privacy violations.

Some technological issues form serious practical bounds for Agreement Technologies. For example, the credential infrastructure (e.g. Digital Ids) have not been picked up yet by industry—standard login/password are used instead.

6 Challenges in large-scale open distributed systems

6.1 Integration with other techniques

Models of negotiation, argumentation and trust, as well as semantics techniques, typically need to be deployed in large-scale open distributed systems such as the grid and cloud computing platforms. Moreover, they may need to be integrated with current technological applications such as crowd-sourcing and social networks, which are also deployed in large-scale open distributed systems. This integration presents in itself several technical and conceptual challenges, for example, concerning the automated assessment of interactions before this assessment can be fed into the trust and reputation model used. In the case of social networks, one of the challenges amounts to understanding the factors determining users' trust in other users and their opinions, such as other users' profiles and users' preferences (see Golbeck, 2009).

6.2 Usability

Usability is one of the major challenges to be addressed to foster widespread adoption in large-scale open systems. More generally, the research community working on negotiation, argumentation, trust and semantics will have to convincingly show that their techniques are compatible with a wide range of application requirements, and provide concrete, measurable evidence of the benefits associated to such models and methodologies. Indeed, the community would benefit from a coherent set of benchmarks, pilot applications and demonstrators providing such concrete evidence.

6.3 Deployment

Despite the promise of argumentation-based (cognitive or hybrid) approaches, these approaches have not been deployed in fully edged systems and applications. Future work is needed to transfer these predominantly theoretical works into practice.

6.4 Quantifying benefits

Another important challenge amounts to quantifying the benefits brought about by Agreement Technologies in order to engage industry. The computations involved in argumentation techniques are more complex than other *ad hoc* solutions (as, e.g. the explicit conflict handling rules adopted in policy languages). In order to foster the adoption of argumentation systems in real-world applications, it is essential to address concrete use cases of industrial interest where the benefits of argumentation solutions clearly emerge (especially from the perspective of adopters and end users) and can be directly compared with standard implementations. Some preliminary work along these lines has been carried out in the context of the ARGUGRID project[5], for e-procurement (see Matt *et al.*, 2008). Similar exercises also provide useful feedback to research, for example, by highlighting pragmatic issues (such as usability) that reduce the effectiveness of theoretically appealing techniques or make them hard to include in a particular application context or work environment.

6.5 Online dispute resolution

An important issue when attempting to deploy geographically distributed platforms to support applications is the fragmented nature or even absence of cross-border legislation (see Stanoevska

[5] www.argugrid.eu

et al., 2008). This can result in legal disputes that are hard to mediate and may undermine the confidence of users in the fruitfulness of deploying these platforms.

Thus, one of the biggest challenges for electronic commerce is how to deal with disputes (especially cross-border ones)[6]. Distance, different laws and jurisdiction are all potential obstacles to online business. Online Dispute Resolution (ODR), which is currently going through a process of regulation, is a clear mechanism for building trust in online transactions in situations where relationships are new or where there is a lack of efficient institutions that cover the transactions. The initial assumption that online markets would just need low prices has been shown to be wrong, as low risk transactions are equally relevant. ODR is also important for simple interpersonal dispute resolution as shown by Zeleznikow (2008) or as used in eBay[7]. See Poblet (2008) for recent advances in the field. A preliminary solution to support ODR for contract-regulated interactions, using argumentation, has been proposed in the ARGUGRID project (see Dung & Thang, 2009).

6.6 Engaging industry

Trust plays a similar role in cloud computing as mentioned in Section 2. Indeed, trust permits to assess the reliability of resources (e.g. PaaS, platform as a service, SaaS, software as a service, or IaaS, Infrastructure as a service). Clouds require the signing of SLAs in order for them to operate. One of the challenges is how to allow the interpersonal operation among commercial clouds (e.g. Amazon, Google) and non-commercial ones. There is room for semantics to help with this interoperability problem. Negotiation and argumentation techniques are going to be fundamental for the settlement of cloud computing agreements.

Some doubts have been shed on the willingness of industry to adopt the cloud computing paradigm[8] and debates between detractors and supporters abound. Some of the arguments against cloud computing are poor security, portability and reliability. Agreement technologies can help meet these challenges. Indeed, semantics techniques can help on improving *portability* by supporting interoperability among platforms and by helping in the translation among formats. Trust and reputation techniques, being social-based decision-making methods, can help on improving *security* and *reliability* by providing the means to select the most trustworthy sites, thus complementing hard security techniques.

7 Conclusions

Agreement technologies show promise in settings where autonomous agents can suitably support humans to identify mutually acceptable and profitable agreements. In this paper, we have considered four specific Agreement Technologies, namely trust, argumentation, negotiation and semantic alignment, and in particular the integration of the former with the latter three. The focus has been on open questions and challenges, building up on current achievements. We have looked at open challenges in general as well as in a specific class of applications, arising in large-scale open distributed systems.

We have ignored other Agreement Technologies, notably the use of *norms* to regulate interactions and the possibility of agents to team up in *organizations*. Naturally, there are some interesting open challenges also concerning the integration of trust with these additional technologies. For example, trust may need to be applied in the context of norms (see Luck *et al.*, 2004), and trust is important in determining the composition of organizations during their formation and evolution (see Young, 2008; McGinnis *et al.*, 2011). Moreover, we have focused on the integration of *trust* with other Agreement

[6] United Nations, E-Commerce and Development Report 2003, `http://www.unctad.org/ecommerce/2003`

[7] `http://pages.ebay.com/services/buyandsell/disputeres.html`

[8] `http://gigaom.com/2008/07/01/10-reasons-enterprises-arent-ready-to-trust-the-cloud/`

Technologies. However, it would be interesting to consider the integration of any Agreement Technologies with any other Agreement Technologies, for example argumentation and norms (see Heras *et al.*, 2009) and norms and organizations (see Argente *et al.*, 2008).

We have left out of this paper those approaches to agreement, in the form of collective decision making, based on social choice (e.g. see Colomer, 2011). Voting and ranking, for instance, are very efficient mechanisms and have been used since the middle ages (see Fidora & Sierra, 2011) to aggregate preferences but lack the fundamental aspect of being the result of a *dialogue* between the parties. The dialogical aspect is fundamental in our view of Agreement Technologies. Dialogues are used to efficiently explore the space of solutions and to select one of them.

As the open challenges are being tackled, the specific combinations of Agreement Technologies involved shall be assessed with suitable demonstrators and systematic benchmarking. There are several computational trust and reputation models that can be put into action and compared along several dimensions when facing realistic situations in open, distributed settings. Specific B2B operations (e.g. in the automotive, textile or oil industries (see Grabowski & Roberts, 1999), disaster rescue (see Camarinha-Matos *et al.*, 2005) as well as particular social networks (e.g. for health care) provide scenario of industrial interest, suitable for the benchmarking and validation of Agreement Technologies and models.

Acknowledgements

This work was partially supported by the Agreement Technology COST action (IC0801). The authors would like to thank for helpful discussions and comments all participants in the panel on 'Trust, Argumentation and Semantics' on 16 December 2009, Agia Napa, Cyprus.

References

Argente, E., Criado, N., Julián, V. & Botti, V. J. 2008. Designing norms in virtual organizations. In *Artificial Intelligence Research and Development, Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence, CCIA 2008, October 22–24, 2008, Sant Mart d'Empúries, Spain, volume 184 of Frontiers in Artificial Intelligence and Applications*, Alsinet, T., Puyol-Gruart, J. & Torras, C. (eds). IOS Press, 16–23.

Baselice, S., Bonatti, P. A. & Faella, M. 2007. On interoperable trust negotiation strategies. In *POLICY*, Agrawal, D., Bertino, E. & Kagal, L. (eds). IEEE Computer Society, 39–50.

Besnard, P. & Hunter, A. 2008. *Elements of Argumentation*. MIT Press.

Bonatti, P. A., De Coi, J. L., Olmedilla, D. & Sauro, L. 2009. Rule-based policy representations and reasoning. In *REWERSE*, Bry, F. & Maluszynski, J. (eds). Lecture Notes in Computer Science **5500**, 201–232. Springer.

Bonatti, P. A., De Coi, J. L., Olmedilla, D. & Sauro, L. 2010. A rule-based trust negotiation system. *IEEE Transactions on Knowledge and Data Engineering* **22**(11), 1507–1520.

Bonatti, P. A., Faella, M., Galdi, C. & Sauro, L. 2011. Towards a mechanism for incentivating privacy. In *Proceedings of the Computer Security – ESORICS 2011 – 16th European Symposium on Research in Computer Security, Leuven, Belgium, September 12–14, 2011*, Atluri V. & Díaz C. (eds). Lecture Notes in Computer Science **6879**, 472–488. Springer.

Bonatti, P. A. & Olmedilla, D. 2005. Driving and monitoring provisional trust negotiation with metapolicies. In *POLICY*, Winsborough, W. & Sahai, A. (eds). IEEE Computer Society, 14–23.

Bonatti, P. A. & Samarati, P. 2003. Logics for authorizations and security. In *Logics for Emerging Applications of Databases*, Chomicki, J., van der Meyden, R. & Saake, G. (eds). Springer, 277–323.

Camarinha-Matos, L. M., Silveri, I., Afsarmanesh, I. H. & Oliveira, A. 2005. Towards a framework for creation of dynamic virtual organizations. In *Collaborative Networks and their Breeding Environments*, Camarinha-Matos, L. M., Afsarmanesh, H. & Ortiz, A. (eds). Springer, 69–80.

Cardoso, H. L. & Oliveira, E. 2005. Virtual enterprise normative framework within electronic institutions. In *Engineering Societies in the Agents World V*, Omicini, A., Gleizes, M. P. & Zambonelli, F. (eds), Springer, 14–32.

Cardoso, H. L. & Oliveira, E. 2008a. Norm defeasibility in an institutional normative framework. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008)*, Ghallab, M., Spyropoulos, C., Fakotakis, N. & Avouris, N. (eds), IOS Press, 468–472.

Cardoso, H. L. & Oliveira, E. 2008b. A context-based institutional normative environment. In *Proceedings of The AAMAS08 Workshop on Coordination, Organization, Institutions and Norms in Agent Systems (COIN)*, Hubner, J., Matson, E., Boissier, E., & Dignum, V. (eds), Springer, 119–133.

Casare, S. J. & Sichman, J. S. 2005. Towards a functional ontology of reputation. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems*, 505–511. Utrecht, The Netherlands.

Castelfranchi, C. & Falcone, R. 2000. Trust is much more than subjective probability: mental components and sources of trust. In *33rd Annual Hawaii International Conference on System Sciences (HICSS-33)*, 4–7 January, 2000, Maui, Hawaii, *Track 6: Internet and the Digital Economy*. IEEE Computer Society.

Choi, N., Song, I.-Y. & Han, H. 2006. A survey on ontology mapping. *SIGMOD Record* **35**(3), 34–41.

Colomer, J. M. 2011. *Social Choice Theory*, International Encyclopedia of Political Science edition. Sage.

Conte, R. & Paolucci, M. 2002. *Reputation in Artificial Societies: Social Beliefs for Social Order*. Kluwer Academic Publishers.

Craswell, R. 2000. Contract law: general theories. In *Encyclopedia of Law and Economics, Volume III: The Regulation of Contracts*, Bouckaert, B. & De Geest, G. (eds). Edward Elgar, 1–24.

Debenham, J. & Sierra, C. 2006. Trust and honour in information-based agency. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi Agent Systems, AAMAS 2006*, 1225–1232. Hakodate, Japan.

Dondio, P. & Barrett, S. 2007. Presumptive selection of trust evidences. In *6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, Honolulu, Hawaii, USA, May 14–18, 2007, Durfee, E. H., Yokoo, M., Huhns, M. N. & Shehory, O. (eds). IFAAMAS, 1078–1085.

Dung, P. M. & Thang, P. M. 2009. Modular argumentation for modelling legal doctrines in common law of contract. *Artificial Intelligence and Law* **17**(3), 167–182.

Fabregues, A. & Sierra, C. 2009. *A Testbed for Multiagent Systems*. Technical report, IIIA-CSIC.

Fidora, A. & Sierra, C. (eds). 2011. *Ramon Llull: From the Ars Magna to Artificial Intelligence*. IIIA-CSIC.

Fullam, K. K., Klos, T. B., Muller, G., Sabater, J., Schlosser, A., Topol, Z., Barber, K. S., Rosenschein, J. S., Vercouter, L. & Voss, M. 2005. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2005)*, 512–518. Utrecht, The Netherlands.

Golbeck, J. 2009. Trust and nuanced profile similarity in online social networks. *ACM Transactions on the Web* **3**(4), 1–33.

Gong, L. & Ellison, G. 2003. *Inside Java(TM) 2 Platform Security: Architecture, API Design, and Implementation*, 2nd edn. Pearson Education.

Grabowski, M. & Roberts, K. H. 1999. Risk mitigation in virtual organizations. *Organization Science* **10**(6), 704–721.

Heras, S., Criado, N., Argente, E. & Julian, V. 2009. Norm emergency through argumentation. *Journal of Physical Agents* **3**, 31–38.

Hogan, A., Harth, A. & Polleres, A. 2009. Scalable authoritative OWL reasoning for the web. *International Journal on Semantic Web and Information Systems* **5**(2), 49–90.

Huynh, T. D., Jennings, N. R. & Shadbolt, N. R. 2006. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems* **13**(2), 119–154.

Jajodia, S., Samarati, P., Sapino, M. L. & Subrahmanian, V. S. 2001. Flexible support for multiple access control policies. *ACM Transactions on Database Systems* **26**(2), 214–260.

Kaplow, L. 2000. General characteristics of rules. In *Encyclopedia of Law and Economics*, **V**: The Economics of Crime and Litigation, Bouckaert, B. & De Geest, G. (eds). Edward Elgar, 502–528.

Karlins, M. & Abelson, H. I. 1970. *Persuasion*. Crosby Lockwood & Son.

Koster, A., Sabater-Mir, J. & Schorlemmer, M. 2010. Inductively generated trust alignments based on shared interactions. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi Agent Systems*, Toronto, Canada.

Krause, P., Ambler, S., Elvang-Gøransson, M. & Fox, J. 1995. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence* **11**, 113–131.

Lacomme, L., Demazeau, Y. & Camps, V. 2009. Personalization of a trust network. In *Agents and Artificial Intelligence*, Filipe, J., Fred, A. & Sharp, B. (eds), Communications in Computer and Information Science, Springer, 247–259.

Lee, A. J. & Winslett, M. 2006. Virtual fingerprinting as a foundation for reputation in open systems. In *iTrust*, Stølen, K., Winsborough, W. H., Martinelli, F. & Massacci F. (eds). Lecture Notes in Computer Science **3986**, 236–251. Springer.

Lenzini, G., Sahli, N. & Eertink, H. 2008. Agents selecting trustworthy recommendations in mobile virtual communities. In *AAMAS-TRUST*, Falcone, R., Suzanne Barber, K., Sabater-Mir, J. & Singh, M. P. (eds). Lecture Notes in Computer Science **5396**, Springer, 182–204.

Luck, M., Munroe, S., Ashri, R. & López y López, F. 2004. Trust and norms for interaction. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*. IEEE, 1944–1949.

Matt, P.-A., Morge, M. & Toni, F. 2010. Combining statistics and arguments to compute trust. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, W. & Kaminka, G. A. (eds). IFAAMAS, Toronto, Canada, 209–216.

Matt, P.-A., Toni, F., Stournaras, T. & Dimitrelos, D. 2008. Argumentation-based agents for e-procurement. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*. Estoril, Portugal.

McGinnis, J., Stathis, K. & Toni, F. 2011. A formal model of agent-oriented virtual organisations and their formation. *Multiagent and Grid Systems* **7**(6), 291–310.

Melaye, D. & Demazeau, Y. 2005. Bayesian dynamic trust model. In Pechoucek, M., Petta, P. & Varga, L. (eds). Multi-Agent Systems and Applications IV, LNCS, **3690**, 480–489. Springer.

Miles, S., Groth, P., Munroe, S. & Moreau, L. 2009a. PrIMe: a methodology for developing provenance-aware applications. *ACM Transactions on Software Engineering and Methodology* **20**(3), 1–42.

Miles, S., Groth, P., Oren, N. & Luck, M. 2009b. Handling mitigating circumstances for electronic contracts. In *Proceedings of the 7th European Workshop on Multi-Agent Systems*, Ayia Napa, Cyprus.

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E. & Van den Bussche, J. 2011. The open provenance model core specification (v1.1). *Future Generation Computer Systems* **27**(6), 743–756.

Mui, L., Mohtashemi, M. & Halberstadt, A. 2002. A computational model for trust and reputation. In *Proceedings of the 35th Hawaii International Conference on System Sciences*, Hawaii, USA.

Noy, N. F. 2004. Semantic integration: a survey of ontology-based approaches. *SIGMOD Record* **33**(4), 65–70.

Ossowski, S. 2008a. Coordination and agreement in multi-agent systems. In *Cooperative Information Agents XII, 12th International Workshop, CIA 2008, Prague, Czech Republic, September 10–12, 2008, Proceedings*, Klusch, M., Pechoucek, M. & Polleres, A., (eds). Lecture Notes in Computer Science **5180**, 16–23. Springer.

Ossowski, S. 2008b. Coordination in multi-agent systems: towards a technology of agreement. In *Multiagent System Technologies, 6th German Conference, MATES 2008*, *Kaiserslautern, Germany, September 23–26, 2008*, *Proceedings*, Bergmann, R., Lindemann, G., Kirn, S. & Pechoucek, M. (eds). Lecture Notes in Computer Science **5244**, 2–12. Springer.

Parsons, S., McBurney, P. & Sklar, E. 2010. Reasoning about trust using argumentation: a position paper. In *Proceedings of the Seventh International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2010), affiliated to AAMAS 2010*, Toronto, Canada.

Pinyol, I. & Sabater-Mir, J. 2009. Towards the de_nition of an argumentation framework using repu-tation information. In *Proceedings of the 12th Workshop on Trust in Agent Societies (TRUST@AAMAS'09)*, Budapest, Hungary.

Pinyol, I., Sabater-Mir, J. & Cuni, G. 2007. How to talk about reputation using a common ontology: from definition to implementation. In *Proceedings of the Ninth Workshop on Trust in Agent Societies*, 90–102, Hawaii, USA.

Poblet, M. (ed.) 2008. *Proceedings of the 5th International Workshop on Online Dispute Resolution, in conjunction with the 21st International Conference on Legal Knowledge and Information Systems (JURIX 2008)*, December 13, 2008, Firenze, Italy, volume 430 of *CEUR Workshop Proceedings*, `CEUR-WS.org`.

Prade, H. & Subrahmanian, V. S. 2007. A qualitative bipolar argumentative view of trust. In *Scalable Uncertainty Management, First International Conference, SUM 2007*, October 10–12, 2007, Washington, DC, USA. Lecture Notes in Computer Science **4772**, 268–276.

Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S. & Sonenberg, L. 2003. Argumentation-based negotiation. *The Knowledge Engineering Review* **18**(4), 343–375.

Rahwan, I. & Simari, G. (ed.). 2009. *Argumentation in AI: The Book*. Springer.

Regan, K. & Cohen, R. 2005. Indirect reputation assessment for adaptive buying agents in electronic markets. *Business Agents and the Semantic Web Workshop* **1**, Victoria, British Columbia, Canada, 121–130.

Sabater, J. & Sierra, C. 2005. Review on computational trust and reputation models. *Artificial Intelligence Review* **24**(1), 33–60.

Sadeh, N. M., Hong, J. I., Cranor, L. F., Fette, I., Kelley, P. G., Prabaker, M. K. & Rao, J. 2009. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing* **13**(6), 401–412.

Sierra, C. & Debenham, J. 2007. The logic negotiation model. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-agent Systems*, 1026–1033, Haway, USA.

Sierra, C. & Debenham, J. 2008. Information-based argumentation. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning: Workshop on Knowledge Representation for Agents and Multiagent Systems (KRAMAS 2008)*, Meyer, J.-J. Ch. & Broersen, J. (eds), 155–170.

Sierra, C. & Debenham, J. 2009. Information-based reputation. In *Proceedings of the First International Conference on Reputation: Theory and Technology*, 5–19, Gargonza, Italy.

Sierra, C. & Sabater-Mir, J. 2005. Review on computational trust and reputation models. *Artificial Intelligence Review* **24**(1), 33–60.

Staab, S., Bhargava, B. K., Lilien, L., Rosenthal, A., Winslett, M., Sloman, M., Dillon, T. S., Chang, E., Hussain, F. K., Nejdl, W., Olmedilla, D. & Kashyap, V. 2004. The pudding of trust. *IEEE Intelligent Systems* **19**(5), 74–88.

Staab, E. & Engel, T. 2008. Combining cognitive with computational trust reasoning. In *Trust in Agent Societies, 11th International Workshop, TRUST 2008*, May 12–13, 2008, Falcone, R., Barber, K. S., Sabater-Mir, J. & Singh, M. P. (eds). Lecture Notes in Computer Science **5396**, Springer.

Stanoevska, K., Parrilli, D. M. & Thanos, G. 2008. BEinGRID: development of business models for the grid industry. In *Grid Economics and Business Models*. Lecture Notes in Computer Science **5206**, 140–151, Springer.

Tavakolifard, M., Herrmann, P. & Ozturk, P. 2009. Analogical trust reasoning. In *Trust Management III*, Ferrari, E., Li, N., Bertino, E. & Karabulut, Y. (eds). Springer, 149–163.

Urbano, J., Rocha, A. P. & Oliveira, E. 2009. Computing confidence values: Does trust dynamics matter? In Lopes L. S., Mariano P., Lau N. & Rocha L. M. (eds). Progress in Artificial Intelligence, Lecture Notes in Artificial Intelligence **5816**, Springer, 520–531.

Urbano, J., Cardoso, H. L. & Oliveira, E. 2010a. Making electronic contracting operational and trustworthy. In *Advances in Artificial Intelligence – Proceedings of the 12th Ibero-American Conference on Artificial Intelligence (IBERAMIA 2010)*, Kuri-Morales, A. & Simari, G. R. (eds). Springer, 264–273.

Urbano, J., Rocha, A. & Oliveira, E. 2010b. Trust estimation using contextual fitness. In *Agent and Multi-Agent Systems: Technologies and Applications*, Jedrzejowicz, P., Nguyen, N., Howlett, R. & Jain, L. (eds). Lecture Notes in Computer Science **6070**, 42–51. Springer.

Urbano, J., Rocha, A. P. & Oliveira, E. 2010c. Trustworthiness tendency incremental extraction using information gain. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010, IEEE/WIC/ACM International Conference*, 2, 411–414.

Urbano, J., Rocha, A. P. & Oliveira, E. 2011. Extracting trustworthiness tendencies using the frequency increase metric. In *Enterprise Information Systems*, Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M. J., Szyperski, C., Filipe, J. & Cordeiro, J. (eds). Lecture Notes in Business Information Processing **73**, 208–221. Springer.

Winslett, M., Yu, T., Seamons, K. E., Hess, A., Jacobson, J., Jarvis, R., Smith, B. & Yu, L. 2002. Negotiating trust on the web. *IEEE Internet Computing* **6**(6), 30–37.

Young, J. 2008. Trust in virtual organisations: a synthesis of the literature. *IJNVO* **5**(3/4), 244–258.

Yu, B. & Singh, M. P. 2002. Distributed reputation management for electronic commerce. *Computational Intelligence* **18**(4), 535–549.

Yu, T., Winslett, M. & Seamons, K. E. 2001. Interoperable strategies in automated trust negotiation. In *CCS '01: Proceedings of the 8th ACM conference on Computer and Communications Security*, Reiter, M. K. & Samarati, P. (eds). ACM Press, 146–155.

Yu, T., Winslett, M. & Seamons, K. 2003. Supporting structured credentials and sensitive policies through interoperable strategies in automated trust negotiation. *ACM Transactions on Information and System Security* **6**(1), 1–42.

Zeleznikow, J. 2008. Beyond interest based bargaining – incorporating interests and fairness in the development of negotiation support systems. In *Proceedings of the 5th International Workshop on Online Dispute Resolution, in conjunction with the 21st International Conference on Legal Knowledge and Information Systems (JURIX 2008), Firenze, Italy, December 13, 2008, volume 430 of CEUR Workshop Proceedings*, Poblet M. (ed.). CEUR-WS.org.