Chapter 25
Argumentation and Trust

Andrew Koster, Jordi Sabater-Mir and Marco Schorlemmer

Abstract In this chapter we discuss the ways in which trust can be combined with argumentation. This is a new field of research that is showing promising approaches to a number of problems in both argumentation and trust. We discuss three ways in which trust and argumentation are combined. The first is to use the trustworthiness of an agent as a level of confidence in the arguments it provides. Parsons et al. take this a step further and consider the ramifications for this in the combination of arguments from different sources. The second way that trust and argumentation are combined is to compute the trustworthiness of an agent based on arguments about its behavior and we discuss two different approaches to this. Finally, argumentation can be used to improve communication about trust. Methods for doing this are discussed at the end of this chapter.

25.1 Introduction

Trust is a technique for dealing with uncertainty regarding other agents' actions and communications. As such, it is a necessary aspect of a reasoning agent in a multiagent system. Such an agent must coordinate and communicate with the other agents in the system. Trust plays a role in deciding with whom to cooperate and which information sources are reliable. However, it is not the only factor in such decisions. Recently work has been done to combine trust in agents' reasoning systems, specif-

Andrew Koster
IIIA - CSIC, Campus UAB, 08193 Bellaterra, Catalonia, Spain, e-mail: andrew@iiia.csic.es

Jordi Sabater-Mir
IIIA - CSIC, Campus UAB, 08193 Bellaterra, Catalonia, Spain, e-mail: jsabater@iiia.csic.es

Marco Schorlemmer
IIIA - CSIC, Campus UAB, 08193 Bellaterra, Catalonia, Spain, e-mail: marco@iiia.csic.es

ically systems that use argumentation, in a number of different manners. The first is by using the trustworthiness of a source of information within an argument to decide whether it is acceptable or not, which we discuss in Section 25.2. The second, discussed in Section 25.3 is to incorporate information from the argumentation into the computation of that same trustworthiness. Finally, argumentation has been used as a method for communicating more accurately about trust, and we describe this in Section 25.4.

25.2 Trusted arguments

One of the problems encountered in a multiagent society is that agents use information from a variety of sources in their reasoning process. Such sources may be more, or less, reliable. Argumentation frameworks [12] provide a way of reasoning using such information, by giving a formal method for dealing with conflicting information, often with degrees of uncertainty. When considering the sources' trustworthiness as a measure of confidence in the information they provide the link between argumentation and trust is obvious. This is precisely the approach taken by Tang et al. [14]. Their work uses the trustworthiness of the information sources as a measure of the probability that information is true.

Parsons et al. abstract from this work and give a formal account of the properties of an argumentation framework when considering different ways of calculating trust and combining arguments [7]. Specifically they try to satisfy the condition:

> If an agent has two arguments A_1 and A_2 where the supports have corresponding sets of agents Ag_1 and Ag_2 then A_1 is stronger than A_2 only if the agent considers A_1 to be more trustworthy than A_2 [7].

This condition states that arguments grounded in information from less trustworthy sources will not be able to defeat arguments with grounds from more trustworthy sources. The work then describes some computational methods for treating trust and argumentation that satisfy this condition. Unfortunately these methods have very strict properties, the most troublesome is the assumption that trust is transitive, while current sociological research indicates that this is only true in very specific situations [1]. Despite this, the work lays a solid theoretical foundation for incorporating trust into argumentation.

Another approach to incorporating trust into argumentation is taken by Villata et al. [15]. Their work takes a similar approach to Tang et al.'s work and explicitly represents the sources providing the different arguments. The major contribution is in allowing argumentation about the trustworthiness of sources. It allows meta-arguments to support, and attack, statements about the trustworthiness of sources. The effect of this argumentation is to change the confidence agents have in the sources providing different arguments, which, in turn, changes the strength of the various arguments. This is thus a combination of two different forms of combining trust and argumentation. In meta-argumentation arguments are used to evaluate

the trustworthiness of agents. In turn this trustworthiness is used as the strength of another set of arguments. This combination seems very powerful, but in relying purely on argumentation for evaluating the trustworthiness, a very coarse concept of trustworthiness is used. As they themselves state:

> Trust is represented *by default* as the absence of an attack towards the sources, or towards the information items and as the presence of evidence in favour of pieces of information [15].

However, trust is a far more complex relationship than this. Trust is a decision, based on, often conflicting, pieces of information, which is why contemporary trust models do not use a binary value for trustworthiness, but rather use a more fine-grained approach, such as a probability that the target will act in a trustworthy manner, or even a probability distribution over the possible outcomes. In the next section we discuss some methods for incorporating argumentation into a statistical model of trust.

25.3 Argument-supported Trust

Prade's model [11] was, insofar as we know, the first model to incorporate argumentation into a trust model. In this work, trust is considered along a variety of dimensions. Specifically, trust is split into trust categories, which represent different behavioural aspects of the target. Each behavioural aspect may be qualified as good, or not good, for a target agent. The trust model consists principally of a rule-base in which levels of trust are related to the target's behaviours. The trust model then uses the target's actual behaviour to perform abduction and find the range in which the trust evaluations must fall. This range is the trust evaluation of a target.

The arguments in Prade's work thus constitute the trust model itself. By performing the abduction with the rules in the trust model, the agent constructs arguments for its observations. The arguments are thus not part of the input of the trust model, but an inherent part of the calculation process. Matt et al. do consider arguments as a separate source of information for calculating the trustworthiness of a target [6].

Matt et al. propose a method for combining justified claims about a target with statistical evidence for that target's behaviour. These justified claims provide context-specific information about an agent's behaviour. The basis for their trust model is Yu & Singh's model [16], which uses a Dempster-Shafer belief function to provide an estimate of whether an agent will fulfill its obligations, given some evidence about that agent's past behaviour. Matt et al. propose to extend this model with a method for evaluating arguments drawn from contracts, in which an agent's obligations are fixed and guarantees are provided about the quality of interactions. Specifically these contracts specify the requirements along a number of dimensions. These dimensions are aspects of an interaction, such as availability, security or reliability. For each dimension an agent wishes to take into account when evaluating trust, it can construct an argument forecasting the target's behaviour with regards to

that dimension, given the specification of a contract. For each dimension d, Matt et al. can construct the following arguments:

- an argument forecasting untrustworthy behaviour, based on the fact that the contract does not provide any guarantee regarding d.
- an argument forecasting trustworthy behaviour, based on the fact that there is a contract guaranteeing a suitable quality of service along dimension d.
- an argument that mitigates a forecasting argument of the second type, on the grounds that the target has, in the past, "most often" violated its contract clauses concerning d.

They then integrate these arguments into Yu & Singh's trust model, by providing new argumentation-based belief functions that combine the information from forecast arguments with evidence. By incorporating more information, the agent should be able to obtain more accurate trust evaluations and Matt et al. show this empirically.

All the methods discussed so far highlight the different aspects of argumentation and trust for dealing with uncertain information; either by applying trust to argumentation in order to get more accurate arguments, or by applying argumentation to trust to obtain more accurate trust evaluations. However there is another useful way to combine trust and argumentation that has not been discussed so far. Evaluating trust often requires communication, but this communication may be unreliable, simply because trust is a subjective concept. By having agents argue about the trust evaluations themselves, an agent may discover whether the other's communicated trust evaluation is useful to it, or whether its interpretation of the various criteria for evaluating trustworthiness are too different from its own criteria [9]. Furthermore, this communication can be used to adapt its own trust model in order to accept more information. Both of these methods are discussed in the next section.

25.4 Arguments about Trust

Trust is a relationship in which, given a certain context, an agent trusts a target to perform a task, resulting in a specific goal being achieved. This context is represented by an agent's beliefs about the environment and the goal is something the trustor wishes to achieve. Therefore trust is an agent's personal and subjective evaluation of a target. When communicating such a subjective evaluation it is often unclear how useful this evaluation is to the receiving agent: it needs to discover whether the context in which the communicated evaluation was made similar to the context in which the receiver needs to evaluate the target. Pinyol proposes a framework to argue about trust evaluations and decide whether another agent's communicated evaluations can be accepted or not [10].

Pinyol starts by modeling the trust model as an inference relation between sentences in $\mathscr{L}_{Rep}$, a first-order language about trust and reputation [8]. This language is defined by a taxonomy of terms used for describing the process of computing trust,

which is discussed in more detail in Chapter 26, Section 26.4.2 of this book. A trust model is considered as a computational process: given a finite set of inputs, such as beliefs about direct experiences or reputation, it calculates a trust evaluation for a target. The semantics of a computational process can be given by the application of a set of inference rules [3]. Following Koster et al.'s formalization of trust models in a similar manner [4], we define this as follows:

Definition 25.1 (Semantics of a trust model). We say that a set of inference rules $\mathscr{I}$ is a specification of a trust model if, given input Δ and the resulting trust computation δ, we have that $\Delta \vdash_T \delta$, i.e., there exists a finite number of applications of inference rules $\iota \in \mathscr{I}$ by which we may infer δ from Δ.

The inference rules themselves depend on the specifics of the computational process and thus the actual trust model being used, but for any computational trust model, such an inference relation exists. For instance, a trust model might have a rule:

$$\frac{img(T,X), rep(T,Y)}{trust(T, \frac{X+Y}{2})}$$

With *img*, *rep* and *trust* predicate symbols in $\mathscr{L}_{Rep}$. For a specific target *Jim*, an agent knows $\{img(Jim,3),\ rep(Jim,5)\}$. It can thus infer $trust(Jim,4)$ using the rule above. For a full example of representing a trust model in inference rules, we refer to [9].

25.4.1 Reasons for having a trust evaluation

Arguments are sentences in the $\mathscr{L}_{Arg}$ language. This language is defined over $\mathscr{L}_{Rep}$. A sentence in $\mathscr{L}_{Arg}$ is a formula $(\Phi : \alpha)$ with $\alpha \in \mathscr{L}_{Rep}$ and $\Phi \subseteq \mathscr{L}_{Rep}$. This definition is based on the framework for defeasible reasoning through argumentation, given by Chesñevar and Simari [2]. Intuitively Φ is the defeasible knowledge required to deduce α. Defeasible knowledge is the knowledge that is rationally compelling, but not deductively valid. The meaning here, is that using the defeasible knowledge Φ and a number of deduction rules, we can deduce α. The defeasible knowledge is introduced in a set of elementary argumentative formulas. These are called *basic declarative units*.

Definition 25.2 (Basic Declarative Units). A basic declarative unit (bdu) is a formula $(\{\alpha\} : \alpha) \in \mathscr{L}_{Arg}$. Additionally, we define an argumentative theory as being a finite set of bdus.

Arguments are constructed using an argumentative theory Γ and the inference relation $\vdash_{Arg}$, characterized by the deduction rules *Intro-BDU, Intro-AND* and *Elim-IMP* from [2]:

Definition 25.3 (Deduction rules of $\mathscr{L}_{Arg}$).

$$\text{Intro-BDU:}\frac{}{(\{\alpha\}:\alpha)}$$

$$\text{Intro-AND:}\frac{(\Phi_1:\alpha_1),\ldots,(\Phi_n:\alpha_n)}{(\bigcup_{i=1}^{n}\Phi_i:\alpha_1\wedge\cdots\wedge\alpha_n)}$$

$$\text{Elim-IMP:}\frac{(\Phi_1:\alpha_1\wedge\cdots\wedge\alpha_n\rightarrow\beta),(\Phi_2:\alpha_1\wedge\cdots\wedge\alpha_n)}{(\Phi_1\cup\Phi_2:\beta)}$$

An argument $(\Phi:\alpha)$ is valid on the basis of an argumentative theory Γ iff $\Gamma\vdash_{Arg}(\Phi:\alpha)$. Because the deduction rules, and thus $\vdash_{Arg}$, are the same for all agents, they can all agree on the validity of such a deduction, however each agent builds its own argumentative theory, using its own trust model. Let $\mathscr{I}$ be the set of inference rules that specify an agent's trust model. Its bdus are generated from a set of $\mathscr{L}_{Rep}$ sentences Δ as follows:

- For any ground element α in Δ, there is a corresponding bdu $(\{\alpha\}:\alpha)$ in $\mathscr{L}_{Arg}$.
- For all $\alpha_1,\ldots,\alpha_n$ such that $\Delta\vdash_T\alpha_k$ for all $k\in[1,n]$, if there exists an application of an inference rule $\iota\in\mathscr{I}$, such that $\frac{\alpha_1,\ldots,\alpha_n}{\beta}$, then there is a bdu:

  $$(\{\alpha_1\wedge\cdots\wedge\alpha_n\rightarrow\beta\}:\alpha_1\wedge\cdots\wedge\alpha_n\rightarrow\beta)$$

  i.e., there is a bdu for every instantiated inference rule for the trust model specified by $\mathscr{I}$.

Continuing the example from above, our agent might have bdus:
$(\{img(Jim,3)\}:\ img(Jim,3))$,
$(\{rep(Jim,5)\}:\ rep(Jim,5))$ and
$(\{img(Jim,3)\ \wedge\ rep(Jim,5)\rightarrow trust(Jim,4)\}:$
$img(Jim,3)\ \wedge\ rep(Jim,5)\rightarrow trust(Jim,4))$.
These bdus constitute an argumentative theory, from which $(\Phi:trust(Jim,4))$ can be inferred, with Φ the union of the defeasible knowledge of the argumentative theory. Similarly, working backwards, an agent can build a valid argument supporting a trust evaluation it believes. Moreover, it can communicate this argument. The other agent, upon receiving such an argument can decide whether or not to accept the trust evaluation. By doing so, the agent effectively filters out communicated trust evaluations that do not coincide with its own frame of reference. However, in a complex domain where trust evaluations can be based on many different criteria, agents might reach the point where they filter out too much information. To reduce the amount of information discarded, agents, when sending a trust evaluation, could personalize their trust evaluations to the receiver.

25.4.2 Personalized Trust Recommendations

Koster et al. present a framework for personalized trust recommendations [4] that builds upon the argumentation framework presented by Pinyol, which we described above. This extension of the argumentation allows agents to communicate about more than just the trust evaluations: it allows agents to connect these trust evaluations to their beliefs and goals. The sender can then tailor its trust model to give a trust recommendation tailored to the receiver's goal, or the two agents can argue about their beliefs about the environment. In this manner agents can personalize their trust recommendations to each other.

For this argumentation framework, it is necessary for agents to be able to justify their trust evaluations using their goals and beliefs. In order to do this, Koster et al. rely on AdapTrust [5], an extension of the BDI agent model [13], that specifies a method for connecting the parameters in a computational trust model to the beliefs and goals an agent has. Before discussing how the argumentation framework allows agents to personalize their trust evaluations to one another's needs, we briefly summarize AdapTrust.

25.4.3 AdapTrust

AdapTrust [5] provides an abstract method for specifying how a trust model is dependent on an agent's goals and beliefs. It is an extension of the Beliefs-Desires-Intentions framework for intelligent agents [13].

Computational trust models are, fundamentally, methods of aggregation: they combine and merge data from several different sources into a single value, the trustworthiness of a target. However, trust is a subjective concept: it is dependent on the beliefs an agent has about its environment and the goal it is trying to achieve by selecting a partner, for which it requires the trust evaluation. Luckily most computational trust models come equipped with a way of implementing this dependency: they have parameters that can be used to adjust the behaviour of the trust model. The aim of AdapTrust is not to present another trust model, but to incorporate existing trust models into an intelligent agent. This can be used to deal with the multi-faceted aspects of trust or, as we discuss in this chapter, adapt the trust model to improve communication about trust.

In any computational trust model, there are parameters that represent criteria for evaluating trustworthiness. For instance, many trust models use a parameter to give less importance to old information than new. This is useful if old information can become outdated and thus new information is more accurate than old. However, in a largely static environment this is not the case. The value of this parameter should be adjusted to the dynamicity of the environment. In general, the parameters of the trust model should be influenced by an agent's changing criteria for evaluating trustworthiness in a changing environment.

Priority System

The parameters of a trust model describe the importance of the different criteria for evaluating trustworthiness. However, it is more useful to consider this the other way round: the relative importance between the different criteria define a set of parameters for the trust model. These criteria are directly under an intelligent agent's control, and thus an agent is able to adapt its trust model. AdapTrust describes the specific techniques necessary to do this. The first of these is $\mathscr{L}_{PL}$, a language to describe the relative importance of any two criteria that influence a parameter of the trust model. AdapTrust uses a subset of first-order logic with a family of predicates to define this importance relation, also called a priority ordering. For each parameter p of the trust model, the binary predicates $\succ_p$ and $=_p$ are defined with the expected properties of strict ordering and equality, respectively. The terms of the language are a set of elements representing the criteria that influence how the trust model should work. A Priority System is defined as a satisfiable theory in this language. For instance, consider an eCommerce environment. If an agent uses a weight w to calculate its evaluation of a sale and it finds the price of an item to be more important than its delivery time, it can have the priority $price \succ_w delivery\_time$ in its Priority System.

Priority Rules

The second technique of AdapTrust is to create the link between, on the one hand, an agent's beliefs and goals and, on the other hand, the priority between the different criteria for evaluating trust. This link makes explicit the adaptive process: a change in an agent's beliefs or goals effects a change in the priorities over the criteria, which changes the parameters of the trust model. The connection between the beliefs or goals and the priorities is made through *priority rules*. The priority rules are specified using another first-order language, $\mathscr{L}_{Rules}$, with predicates $\rightsquigarrow_{Belief}$ and $\rightsquigarrow_{Goal}$ specifying how a set of beliefs, or a goal, respectively, leads to a specific priority relation between two criteria. By using these rules, the priorities are changed when the belief base changes. Additionally this is how the multi-faceted aspect of trust is emphasized: the goal the agent is trying to achieve influences the priority system and thus the trust model. For instance, in the eCommerce example above, our agent might need to buy a bicycle urgently. It then has the goal $buy\_urgent(bicycle)$. For this goal, delivery time is more important than the price, so it has the priority rule $buy\_urgent(bicycle) \rightsquigarrow_{Goal} (delivery\_time \succ_w price)$ and therewith *adapts* its trust model to the requirements of the goal.

We do not go into detail on how these priority rules come to be. They can be programmed by a designer, or generated dynamically by a machine learning algorithm. In this chapter we focus specifically on another method, namely that they can be incorporated through communication with another agent.

25.4.4 Personalizing Trust Recommendations

The argumentation framework by Pinyol et al. that we described earlier in this section does not allow us to completely address the question of what criteria play a role in computing a trust evaluation, let alone connect these to underlying beliefs and goals. AdapTrust can answer this, but does not provide a language in which to do so. In [4], Pinyol et al.'s argumentation framework is extended with concepts from AdapTrust and we summarize that work here.

The priorities that define the trust model's parameters can be incorporated into the argumentative theory. For this, the dependency of the trust model on the beliefs and goal of an agent must be represented in $\mathscr{L}_{Arg}$. In $\mathscr{L}_{Rep}$, the inference rules $\mathscr{I}$ specify a trust model algorithm. However, in AdapTrust this algorithm has parameters that depend on the agent's beliefs and goal. The inference rules should reflect this. The proposed extension of the language is therefore quite straightforward. Rather than using $\mathscr{L}_{Rep}$ as the single language on which the argumentation framework is built, the agent can argue about concepts in $\mathscr{L}_{KR} = \mathscr{L}_{Rep} \cup \mathscr{L}_{PL} \cup \mathscr{L}_{Rules} \cup \mathscr{L}_{Bel} \cup \mathscr{L}_{Goal}$, where $\mathscr{L}_{PL}$ and $\mathscr{L}_{Rules}$ are the languages of the priorities and priority rules, respectively, in AdapTrust, $\mathscr{L}_{Bel}$ the language of the agent's beliefs and $\mathscr{L}_{Goal}$ that of the agent's goals. Let $\Delta \subseteq \mathscr{L}_{Rep}$ and $\delta \in \mathscr{L}_{Rep}$, such that $\Delta \vdash_T \delta$. From Definition 25.1 we know there is a proof applying a finite number of inference rules $\iota \in \mathscr{I}$ for deducing δ from Δ. However, this deduction in AdapTrust depends on a set of the parameters, which we denote *Params*. Therefore, the inference rules must also depend on these parameters. For each $\iota \in \mathscr{I}$, we have $Params_\iota \subseteq Params$, the (possibly empty) subset of parameters corresponding to the inference rule. Let the beliefs Ψ and goal γ determine the values for all these parameters. We denote this as $\Delta \vdash_T^{\Psi,\gamma} \delta$, which states that the trust model infers δ from Δ, given beliefs Ψ and goal γ. Similarly we have $\iota^{\Psi,\gamma} \in \mathscr{I}^{\Psi,\gamma}$ to denote a specific instantiation of the parameters $Params_\iota$ using beliefs Ψ and goal γ.

This allows us to redefine the set of bdus and thus the argumentative theory in such a way that the argumentation supporting a trust evaluation can be followed all the way down to the agent's beliefs and goal. The deduction rules are the same as in Pinyol et al.'s framework, but the bdus for $\mathscr{L}_{Arg}$ are defined as follows in [4]:

Definition 25.4 (Basic Declarative Units for $\mathscr{L}_{Arg}$). Let $\delta \in \mathscr{L}_{Rep}$ be an agent's trust evaluation based on inference rules $\mathscr{I}^{\Psi,\gamma}$, such that $\Delta \vdash_T^{\Psi,\gamma} \delta$ with $\Delta \subseteq \mathscr{L}_{Rep}$, $\Psi \subseteq \mathscr{L}_{Bel}$ and $\gamma \in \mathscr{L}_{Goal}$. For each $\iota \in \mathscr{I}^{\Psi,\gamma}$, let $Params_\iota$ be the corresponding sets of parameters. Let labels be a function that, given a set of parameters, returns a set of constants in $\mathscr{L}_{PL}$, the language of the priority system. Additionally let $\Xi \subseteq \mathscr{L}_{Rules}$ be the agent's set of trust priority rules and $\Pi \subseteq \mathscr{L}_{PL}$ be its priority system based on Ψ and γ, then:

1. For any sentence $\psi \in \Psi$, there is a corresponding bdu $(\{\psi\} : \psi)$ in $\mathscr{L}_{Arg}$.
2. The goal γ has a corresponding bdu $(\{\gamma\} : \gamma)$ in $\mathscr{L}_{Arg}$
3. For all priorities $\pi \in \Pi$ and all the rules $\xi \in \Xi$ the following bdus are generated:

- if ξ has the form $\Phi \rightsquigarrow_{Belief} \pi$ and $\Phi \subseteq \Psi$ then $(\{(\bigwedge_{\varphi \in \Phi} \varphi) \rightarrow \pi\} : (\bigwedge_{\varphi \in \Phi} \varphi) \rightarrow \pi)$ is a bdu in $\mathscr{L}_{Arg}$
- if ξ has the form $\gamma \rightsquigarrow_{Goal} \pi$ then $(\{\gamma \rightarrow \pi\} : \gamma \rightarrow \pi)$ is a bdu in $\mathscr{L}_{Arg}$

4. For all $\alpha_1, \ldots, \alpha_n$ such that $\Delta \vdash_T^{\Phi,\gamma} \alpha_k$ for all $k \in [1,n]$, if there exists an application of an inference rule $\iota^{\Psi,\gamma} \in \mathscr{I}^{\Psi,\gamma}$, such that $\frac{\alpha_1,\ldots,\alpha_n}{\beta}$ and $\mathsf{labels}(Params_{\iota^{\Psi,\gamma}}) = L$ then $(\{(\bigwedge_{\pi \in \Pi_L} \pi) \rightarrow (\alpha_1 \wedge \cdots \wedge \alpha_n \rightarrow \beta)\} : (\bigwedge_{\pi \in \Pi_L} \pi) \rightarrow (\alpha_1 \wedge \cdots \wedge \alpha_n \rightarrow \beta))$ is a bdu of $\mathscr{L}_{Arg}$. With $\Pi_L \subseteq \Pi$ the set of priorities corresponding to labels L.

In items 1 and 2 the relevant elements of the agent's reasoning are added to the argumentation language. In items 3 and 4 the implements for reasoning about trust are added: in 3 the trust priority rules and in 4 the rules of the trust model. The bdus added in 4 contain a double implication: they state that if an agent has the priorities in Π_L then a *trust rule* (which was a bdu in Pinyol's argumentative theory) holds. In practice what this accomplishes, is to allow the argumentation to go a level deeper: agents can now argue about *why* a trust rule, representing an application of a deduction rule in the trust model, holds. An argument for a trust evaluation can be represented in a tree. At each level, a node can be deduced by using the deduction rules of $\mathscr{L}_{Arg}$ with as preconditions the node's children. A leaf in the tree is a bdu. Each agent can construct its own argumentation tree for a trust evaluation and used in a dialogue to communicate personalized trust evaluations. The dialogue starts as an information-seeking dialogue, but if the agents discover their priorities are incompatible, they can discover whether this is due to a lack of information of either agent, or whether their world views are simply incompatible. If either agent is lacking information or the agents think they can reach an agreement on beliefs, they can enter a persuasion dialogue to achieve an agreement on the beliefs and trust priority rules. If this succeeds, they can restart the dialogue and see if they now agree on trust evaluations. In this way the argument serves to allow cooperative agents to converge on a similar model of trust and supply each other with personalized trust recommendations.

25.5 Conclusions

In this section we have given an overview of the ways in which argumentation is used in trust and reputation models and vice versa. We have discussed the application of trust metrics in argumentation frameworks for evaluating the strength of an argument, using the trustworthiness of its information sources. Similarly we have seen how arguments can support trust in various manners. Argumentation about contracts can supply valuable information about an agent's behaviour. Villata et al. [15] combine both types and allow arguments to support trust evaluations in a meta-argument, which in turn decides the strength of an argument at the normal level of argumentation. Finally we discuss two ways in which argumentation can be used in the communication of trust evaluations. The first is a method for deciding whether a communicated trust evaluation is an acceptable source of information. The sec-

ond aims to adapt agents' trust models in order for more sources of information to be acceptable. There is no shortage of productive manners for combining trust and argumentation that is only recently gaining popularity. It is the authors' opinion that both fields can benefit greatly from the tools proposed in the other and we look forward to seeing how the area will develop.

References

1. Castelfranchi, C., Falcone, R.: Trust Theory: A Socio-cognitive and Computational Model. Wiley (2010)
2. Chesñevar, C., Simari, G.: Modelling inference in argumentation through labelled deduction: Formalization and logical properties. Logica Universalis **1**(1), 93–124 (2007)
3. Jones, N.D.: Computability and Complexity: From a Programming Perspective. Foundations of Computing. MIT Press (1997)
4. Koster, A., Sabater-Mir, J., Schorlemmer, M.: Personalizing communication about trust. In: V. Conitzer, M. Winikoff, W. van der Hoek, L. Padgham (eds.) Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12), pp. 517–524. IFAAMAS, Valencia, Spain (2012)
5. Koster, A., Schorlemmer, M., Sabater-Mir, J.: Opening the black box of trust: Reasoning about trust models in a bdi agent. Journal of Logic and Computation (Forthcoming 2012). DOI 10.1093/logcom/EXS003
6. Matt, P.A., Morge, M., Toni, F.: Combining statistics and arguments to compute trust. In: Ninth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'10), pp. 209–216. IFAAMAS, Toronto, Canada (2010)
7. Parsons, S., Tang, Y., Sklar, E., McBurney, P., Cai, K.: Argumentation-basded reasoning in agents with varying degrees of trust. In: Proceedings of AAMAS'11, pp. 879–886. IFAAMAS, Taipei, Taiwan (2011)
8. Pinyol, I., Sabater-Mir, J.: Arguing about reputation. the lrep language. In: A. Artikis, G. O'Hare, K. Stathis, G. Vouros (eds.) Engineering Societies in the Agents World VIII: 8th International Workshop, ESAW 2007, *LNAI*, vol. 4995, pp. 284–299. Springer Verlag (2007)
9. Pinyol, I., Sabater-Mir, J.: Towards the definition of an argumentation framework using reputation information. In: Proceedings of the 12th Workshop on Trust in Agent Societies (TRUST@AAMAS'09), pp. 92–103 (2009)
10. Pinyol Catadau, I.: Milking the reputation cow: Argumentation, reasoning and cognitive agents. Ph.D. thesis, Universitat Politècnica de Catalunya (2010)
11. Prade, H.: A qualitative bipolar argumentative view of trust. In: V. Subrahmnanian, H. Prade (eds.) International Conference on Scalable Uncertainty Management (SUM 2007), *LNAI*, vol. 4772, pp. 268–276. Springer (2007)
12. Rahwan, I., Simari, G.: Argumentation in Artificial Intelligence. Springer (2009)
13. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a bdi-architecture. In: Proc. of KR'91, pp. 473–484. Morgan Kaufmann (1991)
14. Tang, Y., Cai, K., McBurney, P., Parsons, S.: A system of argumentation for reasoning about trust. In: Proceedings of EUMAS'10. Paris, France (2010)
15. Villata, S., Boella, G., Gabbay, D., van der Torre, L.: Arguing about the trustworthiness of information sources. In: W. Liu (ed.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, *LNCS*, vol. 6717, pp. 74–85. Springer (2011)
16. Yu, B., Singh, M.P.: An evidential model of distributed reputation management. In: AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems, pp. 294–301. ACM, New York, NY, USA (2002). DOI http://doi.acm.org/10.1145/544741.544809