Reasoning about Trust using Argumentation: A position paper

Simon Parsons[1,2], Peter McBurney[3], and Elizabeth Sklar[1,2]

[1] Department of Computer & Information Science, Brooklyn College,
City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210 USA
{sklar,parsons}@sci.brooklyn.cuny.edu
[2] Department of Computer Science, Graduate Center
City University of New York, 365 Fifth Avenue, New York, NY 10016, USA
[3] Department of Computer Science, University of Liverpool,
Ashton Building, Ashton Street, Liverpool, L69 3BX, United Kingdom
mcburney@liverpool.ac.uk

Abstract. Trust is a mechanism for managing the uncertainty about autonomous entities and the information they store, and so can play an important role in any decentralized system. As a result, trust has been widely studied in multiagent systems and related fields such as the semantic web. Managing information about trust involves inference with uncertain information, decision making, and dealing with commitments and the provenance of information, all areas to which systems of argumentation have been applied. Here we discuss the application of argumentation to reasoning about trust, identifying some of the components that an argumentation-based system for reasoning about trust would need to contain and sketching the work that would be required to provide such a system.

1 Introduction

Trust is a mechanism for managing the uncertainty about autonomous entities and the information they store. As a result trust can play an important role in any decentralized system. As computer systems have become increasingly distributed, and control in those systems has become more decentralized, trust has steadily become an ever more important concept in computer science.

Trust is an especially important issue from the perspective of autonomous agents and multiagent systems. The premise behind the multiagent systems field is that of developing software agents that will work in the interests of their owners, carrying out their owners' wishes while interacting with other entities. In such interactions, agents will have to reason about the amount that they should trust those other entities, whether they are trusting those entities to carry out some task, or whether they are trusting those entities to not misuse crucial information.

This paper argues that systems of argumentation have an important role to play in reasoning about trust. We start in Section 2 by briefly reviewing work that defines important aspects of trust and giving an extended example which illustrates some of these aspects. Section 3 then briefly reviews some of the work on reasoning about trust and identifies some of the characteristics of any effective system for dealing with trust

information. Building on this discussion, Section 4 then argues that systems of argumentation can handle trust and sketches a specific system of argumentation for doing this. Section 5 then concludes.

2 Trust

As a number of authors have pointed out, trust is a concept that is both complex and rather difficult to pin down precisely and as a result, there are a number of different definitions in the literature. Thus, to pick a few specific examples, Sztompka [26] (cited in [7]) suggests that:

> Trust is a bet about the future contingent actions of others.

while Mcknight and Chervany [20], drawing on a range of existing definitions, define trust as:

> Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible.

and Gambetta [4] states:

> Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends.

While these definitions differ, there are clearly some common elements. There is a degree of uncertainty associated with trust — whether expressed as a subjective probability, as a bet (which, of course, can be expressed as a subjective probability [11]), or as a "feeling of security". Trust is tied up with the relationships between individuals. Trust is related to the actions of individuals and how those actions affect others.

It is also pointed out in a number of places that there are different kinds of trust, what Jøsang et al. [12] call "trust scopes". For example, [12] cites the classification of [9] which identifies the following types of trust:

1. *Provision trust:* the trust that exists between the user of a service or resource, and the provider of that resource.
2. *Access trust:* the trust that exists between the owner of a resource and those that are accessing those resources.
3. *Delegation trust:* the trust that exists between an individual who delegates responsibility for some action or decision and the individual to which that action or decision is delegated.
4. *Identity trust:* trust that an individual is who they claim to be.
5. *Context trust:* trust that an individual has in the existence of sufficient infrastructure to support whatever activities that individual is engaged in.

We illustrate some of these different types of trust with the following example.

Alice is planning a picnic for a group of friends. She asks around amongst some of her aquaintances for ideas about where to hold the picnic. Bob suggests a park a little way outside of the city where he goes quite regularly (provision trust, relating to information) — he says it is quiet and easy to get to. Carol says she has never been to the park herself, but has heard that the bugs are terrible (provision trust, relating to information).

Alice decides that the picnic will be a potluck[4]. Alice asks David to bring potato salad (delegation trust) and Eric says he will bring bread from the bakery near his house (provision trust, relating to a good). Fran offers to bake a cake (provision trust, relating to a good). Carol says she will make her famous barbeque chicken, cooking it on the public barbeques that Alice believes are provided by the park (context trust).

The picnic is scheduled for midday. George arranges to pick up Alice from her house at 10am in order to drive her to the park (Alice doesn't have a car). Harry, who can borrow a minivan (access trust), offers to collect several people from their homes and stop on the way to buy a case of beer. Iain, who is going to ride with George, says he'll bring a soccer ball so they can all play after lunch. John asks if he can bring a friend of a friend, Keith, whom John has never met, but whom John knows will be visiting the city and is unoccupied that day (identity trust).

As Alice makes the arrangements, she is obviously trusting a lot of people to make sure that the plan comes together in ways that are rather distinct.

Bob and Carol are providing information. To decide whether to go to the park, Alice has to factor in the trustworthiness of that information in deciding this, she has to take into account how reliable Bob and Carol are as information providers, not least because the information that they have gven here is contradictory. She might judge that what she knows about Bob (that he goes to the park often) makes him more trustworthy than Carol in this regard (though in other contexts, such as when deciding what film to see, she might value Carol's opinion more), and the fact that Carol is relying on information from yet another person might strengthen this feeling (or, equally, make Alice value Carol's opinion about the park less).

The trust involved in handling the information from Bob and Carol seems to be some what different to the handling of trust when considering the makeup of the meal. Here Alice has to balance not the reliability of the information that people provide, but the *commitments* they are making, the extent to which Carol, David, Eric, Fran, George, Harry and Iain will do what they say they will do. Carol may be a terribly unreliable source of information about parks, and thus untrustworthy in that regard, but a superb provider of barbequed chicken, and one who has never failed to bring that chicken to a potluck when she says that she will. In contrast, Alice may know that Fran saying she will bake a cake means very little. She is just as likely to bake cookies, or realise late the night before the picnic that she has no flour and will have to bring a green salad instead (thus ruining the meal). David, on the other hand, is quite likely not to make

[4] "Pot luck" means that all the guests are expected to bring something that will contribute to the meal, typically an item of food or a beverage.

potato salad; but if he doesn't, he can be relied upon to subsitute it with some close approximation, a pasta salad for example.

In other words, an individual can be an untrustworthy source of information, but a trustworthy provider of services, or indeed an untrustworthy provider of services but a very reliable information source (it is perfectly possible that Fran only ever provides correct information despite her food-related flakiness) — there are different dimensions of trust for different services that are provided (here, information and food items). We distinguish this by talking of the *context* of trust. Similarly, the failure of an individual to fulfill their commitments is not necessarily binary — how they fail can be important.

There are also other aspects to the failure of a commitment. Actions have time and location components. If George is a few minutes late picking Alice up, it may not affect the picnic. If he is an hour late, that might be catastrophic. If he has the wrong address, then even if he arrives at that (wrong) location at 10am, the success of the picnic is in danger. And if Harry can't find his way to the park, there won't be any soccer after lunch even if he successfully collected everyone and bought the beer just as he said he would. However, as long as he arrives while the picnic is going on, then his passengers have a chance to enjoy themselves, though the later he arrives, the less chance that they will have a good time.

3 Reasoning about Trust

As discussed above, a key aspect of trust is that it stems from the relationship between individuals or groups of individuals. This means that it is a relative notion — Alice and Bob may have different views about Carol's trustworthiness — and thus that *provenance* is important in reasoning about trust [6]. A situation that often arises is one where it is necessary to combine different people's information about trust and when this is done, it is important to know where information about trust is coming from.

In this context, Jøsang et al. [12] distinguish between *functional* trust, the trust in an individual to carry out some task, and *referral* trust, the trust in an individual's recommendation. Thus, in our example, Alice's reasoning about George's offer of a lift, and Carol's offer to bring chicken are *functional* trust — Alice is thinking about George's reliability as a provider of lifts and Carol's reliability as a provider of chicken. However, if Alice were to ask Carol for a recommendation for a good butcher, then Alice would base her assessment of Carol's answer on her (Alice's) assesssment of Carol's ability to make good recommendations, an instance of referral trust, while what Carol expresses about her butcher is another instance of functional trust.

As [12] points out, the fact that Carol trusts her butcher to supply good meat is not necessarily a reason for Alice to do the same, and it certainly isn't a reason for Alice to trust the butcher in any more general context (to do a good job of painting Alice's house, for example). However, under certain circumstances — and in particular when the trust context is the same, as it is when Alice is considering the use of Carol's butcher as a provider of meat [14][5] — it is reasonable to consider trust to be transitive. Thus Alice

[5] Depending on the butcher, of course, even this might be too broad a trust context — perhaps the butcher provides excellent chicken and beef, but can only supply indifferent pork and his game has never been hung for long enough.

can consider combining her direct assessment of Carol's referral trustworhiness in the food domain, with Carol's direct assessment of her butcher's functional trustworthiness to derive an *indirect* functional assessment of the butcher.

Given this transitivity, the notion of a *trust network* then makes sense. If Alice can estimate the referral trustworthiness of her friends, and they can do the same for their friends, then Alice can make judgements about recommendations she receives not just from her friends, but also from the friends of her friends (and their friends and so on). The question is, what is a reasonable way to represent this computationally?

At the moment there is no definitive answer to the question. As the definitions of trust cited above suggest, one way to model trust is to use some form of subjective probability — Alice's degree of trust in Bob's park recommendation is a measure of her belief that she will like the park since Bob says that he likes the park. *Eigentrust* [15] is a mechanism, derived for use in peer-to-peer networks, for establishing a global trust rating that estimates how much any individual should trust another. While such a global rating, based as it is on performance, is reasonable for peer-to-peer systems, it has been argued [6] that in the kind of social networks we are discussing here, it is necessary to capture the fact that, for example, Alice and Bob can have very different estimations of Carol's trustworthiness (and, as we have argued, that they will have different ratings for Carol's trustworthiness in different contexts).

Subjective logic [13] is a formalism for capturing exactly this aspect of trust, and for inferring the degree of trust existing between two nodes in a trust network. Based on the Dempster-Shafer theory of evidence [25] it computes a measure that is a generalisation of probability, distinguishing belief in the reliability of an individual, disbelief in the reliability, and the potential belief that has not yet been determined one way or another (termed the "uncertainty"). Singh and colleagues [10, 27] provide extensions of the approach, the former looking at how best to update the measure of trust one individual has in another depending on their experience of interactions. Thus Alice may have her high regard for Carol's food-related recommendations damaged by a bad experience with a supplier that Carol recommends. Subjective logic is not the only approach to handling this problem. For example, Katz and Golbeck [16] describe an algorithm called Tidal-Trust for establishing the trust between a *source* node (representing the individual doing the trusting) and a *sink* mode (representing the individual being trusted). Later work by Kuter and Golbeck provides the SUNNY algorithm [18] which is reported to outperform TidalTrust on a benchmark database of trust information.

4 Argumentation and Trust

The Trust field, including sample literature discussd above, gives us methodologies for *computing* trust, while the Argumentation field can give us methodologies for *reasoning* about trust. In short, we believe that argumentation can provide a mechanism for handling many of the aspects that we need to capture about trust, as we discuss at some length in this section.

4.1 Argumentation in general

As we have discussed above, there are two major aspects that need to be handled by any representation of trust — we need to handle measures of trust, and we need to handle the provenance of trust information. Both of these are provided by several existing argumentation systems.

Some approaches to argumentation, for example abstract approaches such as that of Dung [3] and its derivatives, treat arguments as atomic objects. As a result, they say little or nothing about the internal structure of the argument and have no mechanism to represent the source of the information from which the argument is constructed. Such systems can represent the relationship between arguments ("*a* attacks *b*", and "*b* attacks *c*"), but cannot represent *why* this is the case. As a result, such systems cannot capture the fact that *a* attacks *b* because *b* is based on information from source *s*, and there is evidence that source *s* is not trustworthy.

There are, however, a number of existing systems that extend [3] with more detailed information about the argument. One system system is that of Amgoud [1], where an argument is taken to be a pair (H, h), h being a formula, the *conclusion* of the argument, and H being a set of formulae known as the *grounds* or *support* of the argument. Conclusion and support are related. In particular, [1] requires that H be a minimal consistent set of formulae such that $H \vdash h$ in the language in which h and H are expressed. This means of representing the support is rather restricted. It presents the support as a bag of formulae with no indication as to how they are used in the construction of the argument, and without recording any of the intermediate steps. It is easy enough to see if another argument *rebuts* (H, h), meaning that the conclusion of this second argument is the negation of h, and it is also quite simple to establish if the conclusion of the second argument contradicts any of the grounds in H (which in some systems of argumentation is known as *undercutting*). However, other forms of relationship are harder to establish. For example, in some cases it is interesting to know if an argument contradicts any of the intermediate steps in the chain of inferences between H and h.

Since the information about the steps in the argument can be useful, some systems of argumentation, for example [5] and [22], record more detail about the proof of h from H as part of the grounds. Some, including the system [19] which we will discuss in more detail below, go as far as to record the proof rules used in deriving h from H, permitting the notion of "attack" to include not only the intermediate conclusions but also the means by which they were derived.

Another problem with Dung's argumentation system from the perspective of reasoning about trust is that it has no explicit means to represent degrees of trust. In [3] the important question is whether, given all the arguments that are known, a specific argument should be considered to hold. While one could construct a system for reasoning about trust in this way — the critical point, after all, is often whether someone's argument is trustworthy or not — the prevelance of numerical measures of trust in the literature leads us to want to represent these.

Systems like that of Amgoud [1] provide one means of handling such measures, allowing formulae to have preference values attached to them. The values propagate to arguments and are taken into consideration when reasoning about the relationship between arguments (roughly speaking, strong arguments shrug off the attacks of weaker

arguments). This approach seems a little too restrictive for dealing with trust, but there are systems that are more flexible. One example is the work of Oren et al. [21], which allows formulae and arguments to be weighted with the belief values used by Jøsang's subjective logic [13]. A more abstract approach is that of Fox [17] where values to represent belief in formulae are picked from some suitable *dictionary* of values, and propagated in a suitable way through the proof rules that are used to construct arguments. Arguments are then triples of conclusion, support, and value, and such systems are close to the notion of a *labelled deductive system* [2] (though they pre-date labelled deductive systems by some years).

4.2 A suitable argumentation system

Having given a high level description of how argumentation can help in handling a number of the aspects of reasoning about trust, we give a more detailed example of using a specific system of argumentation. The system we describe is the system *TL* that we introduced in [19], notable because it explicitly represents the rules of inference employed in constructing arguments in the support of the argument (which then makes it possible to dispute the application of those rules).

We start with a set of atomic propositions including $\top$ and $\bot$, the ever true and ever false propositions. The set of well-formed formulae (*wff*s), labeled $\mathcal{L}$, is comprised of the set of atomic propositions closed under the connectives $\{\neg, \rightarrow, \wedge, \vee\}$. $\mathcal{L}$ may then be used to create a database Δ whose elements are 4-tuples:

$$(\theta : G : R : \tilde{d})$$

in which each element θ is a formulae, G is the derivation of that formula, R is the sequence of rules of inference used in the derivation, and $\tilde{d}$ is a suitable measure.

In more detail, θ is a *wff* from $\mathcal{L}$, $G = (\theta_0, \theta_1, \ldots, \theta_{n-1})$ is an ordered sequence of *wff*s, with $n \geq 1$, and $R = (\vdash_1, \vdash_2, \ldots, \vdash_n)$ is an ordered sequence of inference rules, such that:

$$\theta_0 \ \vdash_1 \ \theta_1 \ \vdash_2 \ \theta_2 \ \ldots \ \theta_{n-1} \ \vdash_n \ \theta$$

In other words, each element $\theta_k \in G$ is derived from the preceding element θ_{k-1} as a result of the application of the k-th rule of inference, $\vdash_k$, $(k = 1, \ldots, n-1)$. The rules of inference in any such sequence may be non-distinct. Thus G and R together provide an explicit representation of the way that θ was inferred.

The element $\tilde{d} = (d_1, d_2, \ldots, d_n)$ is an ordered sequence of elements from some dictionary $\mathcal{D}$. For reasoning about trust, these elements could be a numerical measure of trust, or some linguistic term that indicates the trust in the relevant inference, for example:

$$\{\textit{very reliable}, \textit{reliable}, \textit{no opinion}, \textit{somewhat unreliable}, \textit{very unreliable}\}$$

We also permit *wff*s $\theta \in \mathcal{L}$ to be elements of Δ, by including tuples of the form $(\theta : \emptyset : \emptyset : \emptyset)$, where each $\emptyset$ indicates a null term. (Such tuples represent information that has not been derived — basic premises may take this form.) Note that the assignment of labels may be context-dependent, i.e., the d_i assigned to $\vdash_i$ may also depend on θ_{i-1}.

$$\text{Ax}\frac{(\theta : G : R : \tilde{d}) \in \Delta}{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d})}$$

$$\wedge\text{-I}\frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\theta \wedge \phi : G \otimes H \otimes (\theta \wedge \phi) : R \otimes S \otimes (\vdash_{\wedge\text{-I}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\wedge\text{-I}}))}$$

$$\wedge\text{-E1}\frac{\Delta \vdash_{TCR} (\theta \wedge \phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta : G \otimes (\theta) : R \otimes (\vdash_{\wedge\text{-E1}}) : \tilde{d} \otimes (d_{\wedge\text{-E1}}))}$$

$$\wedge\text{-E2}\frac{\Delta \vdash_{TCR} (\theta \wedge \phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\phi : G \otimes (\phi) : R \otimes (\vdash_{\wedge\text{-E2}}) : \tilde{d} \otimes (d_{\wedge\text{-E2}}))}$$

$$\vee\text{-I1}\frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta \vee \phi : G \otimes (\theta \vee \phi) : R \otimes (\vdash_{\vee\text{-I1}}) : \tilde{d} \otimes (d_{\vee\text{-I1}}))}$$

$$\vee\text{-I2}\frac{\Delta \vdash_{TCR} (\phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\theta \vee \phi : H \otimes (\theta \vee \phi) : S \otimes (\vdash_{\vee\text{-I2}}) : \tilde{e} \otimes (e_{\vee\text{-I2}}))}$$

$$\vee\text{-E}\frac{\begin{array}{c}\Delta \vdash_{TCR} (\theta \vee \phi : G : R : \tilde{d}) \text{ and} \\ \Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\gamma : H : S : \tilde{e}) \text{ and } \Delta, (\phi : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\gamma : J : T : \tilde{f}).\end{array}}{\Delta \vdash_{TCR} (\gamma : G \otimes H \otimes J \otimes (\gamma) : R \otimes S \otimes T \otimes (\vdash_{\vee\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes \tilde{f} \otimes (d_{\vee\text{-E}}))}$$

$$\neg\text{-I}\frac{\Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\bot : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\neg\theta : G \otimes (\neg\theta) : R \otimes (\vdash_{\neg\text{-I}}) : \tilde{d} \otimes (d_{\neg\text{-I}}))}$$

$$\neg\text{-E}\frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\neg\theta : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\bot : G \otimes H \otimes (\bot) : R \otimes S \otimes (\vdash_{\neg\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\neg\text{-E}}))}$$

$$\neg\neg\text{-E}\frac{\Delta \vdash_{TCR} (\neg\neg\theta : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta : G \otimes (\theta) : R \otimes (\vdash_{\neg\neg\text{-E}}) : \tilde{d} \otimes (d_{\neg\neg\text{-E}}))}$$

$$\rightarrow\text{-I}\frac{\Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta \rightarrow \phi : G \otimes (\theta \rightarrow \phi) : R \otimes (\vdash_{\rightarrow\text{-I}}) : \tilde{d} \otimes (d_{\rightarrow\text{-I}}))}$$

$$\rightarrow\text{-E}\frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\theta \rightarrow \phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\phi : G \otimes H \otimes (\phi) : R \otimes S \otimes (\vdash_{\rightarrow\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\rightarrow\text{-E}}))}$$

Fig. 1. The TL Consequence Relation

This is the case for statistical inference, where the p-value depends on characteristics of the sample from which the inference is made, such as its size.

With this formal system, we can take a database Δ and use the consequence relation $\vdash_{TCR}$ defined in Figure 1 to build arguments for propositions of interest. This consequence relation is defined in terms of rules for building new arguments from old. The rules are written in a style similar to standard Gentzen proof rules, with the antecedents of the rule above the horizontal line and the consequent below. In Figure 1, we use the notation $G \otimes H$ to refer to that ordered sequence created from appending the elements of sequence H after the elements of sequence G, each in their respective order. The rules are as follows:

Ax The rule Ax says that if the tuple $(\theta : G : R : \tilde{d})$ is in the database, then it is possible to build the argument $(\theta : G : R : \tilde{d})$ from the database. The rule thus allows the construction of arguments from database items.

$\wedge$-I The rule $\wedge$-I says that if the arguments $(\theta : G : R : \tilde{d})$ and $(\phi : H : S : \tilde{e})$ may be built from the database, then an argument for $\theta \wedge \phi$ may also be built. The rule thus shows how to introduce arguments about conjunctions; using it requires an inference of the form: $\theta, \phi \vdash (\theta \wedge \phi)$, which we denote

$$\vdash_{\wedge\text{-I}}$$

in Figure 1. This inference is then assigned a value of $d_{\wedge\text{-I}}$.

$\wedge$-E The rule $\wedge$-E1 says that if it is possible to build an argument for $\theta \wedge \phi$ from the database, then it is also possible to build an argument for θ. Thus the rule allows the elimination of one conjunct from an argument, and its use requires an inference of the form: $\theta \wedge \phi \vdash \theta$. $\wedge$-E2 allows the elimination of the other disjunct.

$\vee$-I The rule $\vee$-I1 allows the introduction of a disjunction from the left disjunct and the rule $\vee$-I2 allows the introduction of a disjunction from the right disjunct.

$\vee$-E The rule $\vee$-E allows the elimination of a disjunction and its replacement by tuple when that tuple is a TL-consequence of each disjunct.

$\neg$-I The rule $\neg$-I allows the introduction of negation.

$\neg$-E The rule $\neg$-E allows the derivation of $\perp$, the ever-false proposition, from a contradiction.

$\neg\neg$-E The rule $\neg\neg$-E allows the elimination of a double negation, and thus permits the assertion of the Law of the Excluded Middle (LEM).

$\rightarrow$-I The rule $\rightarrow$-I says that if on adding a tuple $(\theta : \emptyset : \emptyset : \emptyset)$ to a database, where $\theta \in \mathcal{L}$, it is possible to conclude ϕ, then there is an argument for $\theta \rightarrow \phi$. The rule thus allows the introduction of $\rightarrow$ into arguments.

$\rightarrow$-E The rule $\rightarrow$-E says that from an argument for θ and an argument for $\theta \rightarrow \phi$ it is possible to build an argument for ϕ. The rule thus allows the elimination of $\rightarrow$ from arguments and is analogous to MP in standard propositional logic.

This is an intentionally abstract formalism — syntactically complete, but without a specified semantics. The idea is that to capture a specific domain, we have to identify a suitable dictionary from which to construct the $\tilde{d}$ and that this set of values will determine the mechanism by which we can compute an overall value from the sequence

of d_i. For example, if one wanted to use Jøsang's subjective logic, then the mechanism for combining the d_i's would be taken from [13]. If one wanted to quantify trust using probability, then the combination rules would be those dictated by probability theory (for example using [28]). If one wanted to use the dictionary mentioned above ("very reliable" and so on) then it would be necessary to determine the right way to combine these values across all the inference rules in Figure 1.

Even without specifying these mechanisms, it should be clear that whatever means we use to quantify trust in combination with *TL*, the formalism can both capture trust values and the precise source of information used. It is also possible to go further. The fact that *TL* includes explicit reference to different forms of inference allows us to capture the fact that inferences may differ depending on the source of the information on which they are based — we might want to make different inferences depending on whether the source was something we have direct experience of or something that comes from a trusted source, or something that comes from an untrusted source.

4.3 Extensions

The previous sections have argued that systems of argumentation can provide the core functionality required to reason about trust. Here we discuss how systems of argumentation, especially the system *TL* sketched above, can provide additional mechanisms that are important in dealing with trust.

First, argumentation systems explicitly allow the representation of different points of view. The system *TL* we have sketched above provides us with the rules for constructing arguments, and it does not limit the number of arguments that one can construct for a specific conclusion. Thus, the database Δ may contain information that represents a number of different assessments of the trustworthiness of, for example, a source of information. This might be done through the inclusion of a number of tuples $(\theta : G : R : \tilde{d})$ with different Gs, representing different views of the sources, and different $\tilde{d}$s representing different assessments of trustworthiness. These pieces of information could then be used to make different inferences, with any potential choice between conclusions being made on the basis of the relevant $\tilde{d}$ values.

That is one, fairly simple, way to represent different viewpoints. Another would be to have different argumentation systems represent the views of different individuals, and to use the mechanisms of argumentation-based dialogue (like those discussed in [24, 8]) to explore the differences in the views of trust and to attempt to resolve them. In such a combination, the individual argumentation systems can be constructed using *TL*, and would then reason about trust based on a single viewpoint. The interaction between different viewpoints is then captured by the dialogue mechanisms of [24, 8], enabling a rational discourse about trust issues.

Another important aspect of reasoning about trust, addressed in [10] for example, is the need for an individual to be able to revise the trust they have in another based on experience. Revision of beliefs is not a subject that has been widely considered within the argumentation community, but [23] suggests some approaches to the subject, and these can be implemented on top of *TL*. This would allow us to represent the case in which one individual revises its view of a source as a result of considering information provided by another individual.

5 Conclusion

This paper has presented the case for using argumentation as a mechanism for reasoning about trust. Starting from some of the many views of trust expressed in the literature, we extracted the major features that need to be represented, discussed formalisms for handling trust, and then suggested how argumentation could be used for reasoning about trust. We sketched in some detail how a specific system of argumentation, *TL*, could be used in this way and identified some additional argumentation-based mechanisms that could be of use in dealing with trust.

Acknowledgement

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

1. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.
2. C.Chesñevar and G. Simari. Modelling inference in argumentation through labelled deduction: Formalization and logical properties. *Logica Universalis*, 1(1):93–124, 2007.
3. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
4. D. Gambetta. Can we trust them? In D. Gambetta, editor, *Trust: Making and beraking cooperative relations*, pages 213–238. Blackwell, Oxford, UK, 1990.
5. A. J. Garcia and G. R. Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(2):95–138, 2004.
6. J. Golbeck. Combining provenance with trust in social networks for semantic web content filtering. In *Proceedings of the International Provenance and Annotation Workshop*, Chicago, Illinois, May 2006.
7. J. Golbeck and C. Halaschek-Wiener. Trust-based revision for expressive web syndication. *The Logic Journal of the IGPL*, (to appear).
8. T. F. Gordon. The pleadings game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2(4):239–292, 1994.
9. T. Grandison and M. Sloman. A survey of trust in internet applications. *IEEE Communications Surveys and Tutorials*, 4(4):2–16, 2000.
10. C.-W. Hang, Y. Wang, and M. P. Singh. An adaptive probabilistic trust model and its evaluation. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, 2008.
11. E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

12. A. Jøsang, E. Gray, and M. Kinateder. Simplification and analysis of transitive trust networks. *Web Intelligence and Agent Systems*, 4(2):139–161, 2006.
13. A. Jøsang, R. Hayward, and S. Pope. Trust network analysis with subjective logic. In *Proceedings of the 29th Australasian Computer Society Conference*, Hobart, January 2006.
14. A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.
15. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The Eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th World Wide Web Conference*, May 2004.
16. Y. Katz and J. Golbeck. Social network-based trust in prioritzed default logic. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006.
17. P. Krause, S. Ambler, M. Elvang-Gørannson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.
18. Y. Kuter and J. Golbeck. SUNNY: A new algorithm for trust inference in social networks using probabilistic confidence models. In *Proceedings of the 22nd National Conference on Artificial Intelligence*, 2007.
19. P. McBurney and S. Parsons. Tenacious tortoises: A formalism for argument over rules of inference. In *Proceedings of the ECAI Workshop on Computational Dialectics*, Berlin, 2000.
20. D. H. McKnight and N. L. Chervany. The meanings of trust. Working Paper 96-04, Carlson School of Management, University of Minnesota, 1996.
21. N. Oren, T. Norman, and A. Preece. Subjective logic and arguing with evidence. *Artificial Intelligence*, 171(10–15):838–854, 2007.
22. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261—292, 1998.
23. S. Parsons and E. Sklar. How agents alter their beliefs after an argumentation-based dialogue. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems, Second International Workshop*, volume 4049 of *Lecture Notes in Computer Science*, pages 297–312. Springer, 2005.
24. H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15(6):1009–1040, 2005.
25. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
26. P. Sztompka. *Trust: A Sociological Theory*. Cambridge University Press, Cambridge, UK, 1999.
27. Y. Wang and M. P. Singh. Trust representation and aggregation in a distributed agent system. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006.
28. Y. Xiang and N. Jia. Modeling causal reinforcement and undermining for CPT elicitation. *IEEE Transactions on Knowledge and Data Engineering*, 19(12):1708–1718, 2007.