12

Trust and Technology

In this book our effort is to model and rationalize the trust notion trying to catch all the different and varying aspects of this broad concept. In fact, there are today many studies, models, simulations and experiments trying to integrate trust in the technological infrastructures: The most advanced disciplines in Human-Computer Interaction (Baecker *et al.*, 1987), (Card *et al.*, 1983) and (Dix *et al.*, 2004), Distributed Artificial Intelligence (Lesser, 1990), (Hewitt and Inman, 1991), (Weiss, 1997), Multi-Agent Systems (Wooldridge, 2002), (Shoham and Leyton-Brown, 2008), and Networked-Computer Systems (Grid, Semantic Web, etc. (Foster and Kesselman, 2003), (Antoniou and van Harmelen, 2008) and (Davies, 2006)) are forced to cope with trust.

But why does trust seem to be so important in the advanced technological contexts? Is it necessary to involve such a complex, fuzzy and human related concept? Is it not sufficient to consider just more technical and simply applicable notions like security?

To give a satisfactory response to these questions we have to evaluate which kind of network infrastructures are taken into consideration in the new communication and interaction scenarios, which kind of peculiar features should have the artificial agents we have to cope with, which kind of computing is going to invade (pervade) the future physical environments?

In fact, trust becomes fundamental in the *open* multi-agent systems where the agents (which could be both human beings and artificial agents owned by other human stakeholders) can (more or less freely) enter and leave the system. The evolution of the interaction and communication technological paradigms toward human style, is, on the one hand, a really difficult task to realize, but, on the other hand, it potentially increases the people accessing to (and fruitful in using) the new technologies. In fact, in the history of their evolution humans have learned to cooperate in many ways and environments; on different tasks; and to achieve different goals. They have intentionally realized (or they were spontaneously emerging) diverse cooperative constructs (purely interactional, technical-legal, organizational, socio-cognitive, etc.) for establishing trust among them.

It is now necessary to remodel the trust concept in the new current and future scenarios (new channels and infrastructures of communication; new artificial entities, new environments) and the efforts in the previously cited scientific fields (HCI, MAS, DAI, NCS) are trying to give positive answers to these main requirements.

Trust Theory: A Socio-Cognitive and Computational Model Cristiano Castelfranchi and Rino Falcone
 ISBN: 978-0-470-02875-9

Without establishing trustworthy relationships, these new infrastructures and services, these new artificial agents, these new robots, these new pervasive technologies, do not impact with sufficient strength and in fact do not really integrate with the real society.

One of the main features of these new artificial entities (but in fact of this new technological paradigm) is in particular, *autonomy*: having the capacity to realize tasks without direct human control and monitoring, having the capacity to attribute and permit the single (human or artificial) entities to realize their own goals, having the capacity to make decisions on the basis of their own attitudes, beliefs and evaluations (see Chapter 7 for a detailed analysis of this concept and for its relationships with trust). The new environments are increasing the autonomy levels and complexities of these agents offering sophisticated interaction and cooperation. In these environments no agent can know everything, there is no central authority that controls all the agents (due to the features of the environment).

At the same time these complex autonomies (in open environments, broad communities and with indirect interaction) increase human diffidence and risks.

Technology should not only be reliable, safe, secure, but it *should be also perceived* as such, the user must believe that it is reliable, and must feel confident while using it and depending on it. *The unique real answer for coping with others' autonomy is to establish a real trust relationship.* For these reasons, the ability of understand and model the trust concept to transfer its utility in the technological cooperative framework will be in fact the *bottleneck* of the development of the autonomy-based technology that is the technology of the future.

12.1 Main Difference Between Security and Trust

One important thing to underline is the conceptual difference between the *two notions of security and trust*. In general, a secure system should provide mechanisms (Wong and Sycara, 2000) able to contrast (oppose) potential threats and guarantee a set of features:

- *certainty of identification*: in particular techniques of authentication should be able to identify the interacting agents; this identification allows accessibility to defined rights and resources (Grandison and Sloman, 2000);
- *integrity*: the messages and the actions of the agents should not be corrupted by a third party;
- *confidentiality and not intrusivity*: the communication and interaction should remain private if the decision of the agents is so;
- *nonrepudiation*: in specific cases it should be possible to identify unambiguously the author of messages or actions, and they cannot deny this objective identification;
- *secure delegation*: it should be clear who is the delegator of each agent.

There are various research areas (*encryption* (Ellis and Speed, 2001), *cryptography* (Schneier, 1996), (Stallings, 1999), *authentication* (Stallings, 2001), *access control* (Anderson, 2001)) that develop techniques for achieving the above specified features of security.

The objective of automating the procedures of the traditional security systems has viewed and currently views many studies and applications, some of them make explicit reference to trust even if this concept is used in a very reductive and basic sense, oriented toward the strict security rather than to the more complex and general concept of trust. Examples of this use are the so called *Trusted Systems* (Abrams, 1995); the so called *Trusted Computing* (mainly

used by industry and regarding the information being processed on a platform with specialized security hardware (see (Josang, 2007) for more details); the so called *Trust Management* ((Blaze *et al.*, 1996)) mainly used for security in the distributed access control; the so called *Trusted Third Parthy* (TTP, (Skevington and Hart, 1995)) describing systems in which the presence of a reputed, disinterested, impartial and responsible entity that is accepted by all the parties is guaranted.

The most interesting, recent works (relatively more oriented towards the very concept of trust) in this area are: *Trust-Serv* (Skogsrud *et al.*, 2003), *PolicyMaker* (Grandison and Sloman, 2000), and *KAoS* (Uszok *et al.*, 2004). In these last systems and approaches the main goal is to provide agents with credentials able to obtain trust from the system on the basis of predefined policies. We have to say that even if these systems are a step towards the real concept of trust, in general the main problem of the multi-agent systems is about how an agent can rely on other agents for achieving its own goals (Huynh *et al.*, 2006).

An interesting distinction in the field is described by Rasmussen and Jansson (Rasmussen and Jansson, 1996) between *hard security* and *soft security*, where hard security is referred to the traditional IT (Information Technology) security mechanisms such as those above defined (access control, authentication, and so on) while soft security is about deceitful and malicious service providers that provide misleading, tricky or false information (Rasmussen and Jansson called this security 'social control mechanisms').

In general we can say that *establishing a true trust relationship is a more complex and different thing with respect to security matter*: the above described techniques cannot guarantee that an interaction partner has the competence he claims or that he is honest about his own intentions.

Trust is more than secure communication, e.g., via public key cryptography techniques: the reliability of information about the status of your trade partner has little to do with secure communication or with its identification. *Maybe perceived security and safety are a precondition and also an aspect of trust, but trust is a more complex and broad phenomenon. Trust must give us tools for acting in a world that is in principle insecure where we have to make the decision to rely on someone in risky situations*.

For this reason the trust challenge is more complex and advanced (and therefore more ambitious) than the one about security, even if there are relationships between them and the solutions to the security problems represent a useful basis for coping with trust problems.

12.2 Trust Models and Technology

In the last fifteen years many studies and researches have been developed in the technological field on trust (for a resume see (Marsh, 1994), (Castelfranchi and Tan, 1999), (Falcone *et al.*, 2001), (Falcone *et al.*, 2003), (Ramchurn *et al.*, 2004), (Falcone *et al.*, 2005), (Huynh *et al.*, 2006), (Cofta, 2007), (Falcone *et al.*, 2008), (Golbeck, 2009)). These works have analyzed different aspects, models, and approaches to trust with the common goal of understanding and introducing trusted relationships within the computational framework.

Let us consider the most relevant approaches to trust in the technological domain: the *logical approaches*, the *computational approaches*, and the *socio-cognitive approaches*. These three approaches often have a varying overlap with each other, but the differences are given by the relevance of the goals they mainly are pursuing. In this part of the book we will omit to

introduce the socio-cognitive approaches, which have been discussed at length in the other chapters. In addition you can refer to Chapter 11 for a description of a fuzzy implementation model for the socio-cognitive approach to trust.

12.2.1 Logical Approaches

The *logical approaches* ((Cohen and Levesque, 1990), (Fagin e*t al.*, 1994), (Demolombe, 1999), (Jones and Firozabadi, 2001), (Josang, 2001), (Liau, 2003), (Lorini and Demolombe, 2008), (Castelfranchi *et al.*, 2009)) start from models based on mathematical logics for describing, analyzing and implementing the trust relationships. These approaches have the advantage of using powerful methods able to produce inferences and strongly rationalize the conceptual apparatus. The drawbacks are given as the constraints introduced and derived through the same logics that in fact impose their own rules to the conceptualization of thought and action, very often without considering all the elementary criteria of flexibility of the human reasoning and action. In various cases, the approximation of the formalisms to the reality can satisfy specific descriptive purposes of the reality, in other cases this approximation is not appropriate (in the sense that it does not introduce realistic constraints).

A very elegant example of logical approach is given by the Demolombe's analysis (Demolombe, 1999) with respect to the trust in information sources. With a clear (but functional) simplification (maybe too superficial with respect to the social concepts he intends to model), he considers that each information source can have four different properties: *sincerity*, *credibility*, *cooperativity*, and *vigilance*.

- An agent X is *sincere* with respect to another agent Y and a specific content p if X believes p when he is informing Y about p.
- An agent X is *credible* with respect to another agent Y and a content p if X believes p and p is true in the world.
- An agent X is *cooperative* with respect to another agent Y if what he believes is communicated to Y.
- An agent X is *vigilant* with respect to the world if what is true in the world is believed by X.

Demolombe also derives two other concepts: *Validity*, as the conjunction of sincerity and credibility; and *Completeness*, as the conjunction of cooperativity and vigilance.

Using the modal logic (Chellas, 1990) Demolombe formalizes these concepts and defines different kinds of trust: trust with respect to sincerity, credibility, cooperativity, and vigilance. Then he is able to derive consequences from this representation: For example, he is able to derive additional properties like 'if Y trusts X as sincere, then the information p (received by X) let infer to Y that X believes p', and so on.

Jones (Jones and Firozabadi, 2001) represents some interesting aspects of trust (like deception in terms of trust, trust in other's trust) applying the logic of belief together with the deontic logic, and with the logic of 'count as'. We deeply analyze the Jones's model in another part of this book (see Chapters 2 and 8 and in particular Section 2.2.2).

One of the main problems of applying the classical logical framework to the mental attitudes (on which trust is based) is the difficulty of taking into consideration *uncertainty* and *ignorance*: typical features of the beliefs. Different and interesting attempts (in particular the one of

Dempster and Shafer (Shafer, 1976) to cope with these limits are presented in (Motro and Smets, 1997). However, given his direct work on trust and reputation models, it is of interest to cite Josang's approach (Josang, 2001) that introduces the *subjective logic*: an attempt to overcome the limits of the classical logics and also taking into consideration the uncertainty, ignorance and the subjective characteristic of the beliefs.

This approach is strongly influenced by Dempster and Shafer's work but with some specific interesting intuitions. The *opinions* in Josang's approach are belief/trust metrics denoted by:

$$\omega_x^A = (b, d, u, a) \tag{12.1}$$

where ω_x^A expresses the trust of agent A about the truth of the statement of x. b represents the positive A's beliefs, d represents the A's disbeliefs, u represents the uncertainty with:

$$b, d, u \in [0, 1] \qquad b + d + u = 1 \qquad a \in [0, 1]$$

The parameter a is called *relative atomicity* and represents the base rate probability in the absence of evidence: it determines how uncertainty contributes to an opinion's probability expectation value $E(\omega_x^A)$:

$$E\left(\omega_x^A\right) = b + au \tag{12.2}$$

The subjective logics introduce two operators (*discounting* and *consensus*) useful for trust derivation from other opinions.

12.2.2 Computational Approach

The *computational approach* to trust has as a main goal the implementation of a trust model in an automatic system independent from the representational framework. The computational trust models can follow different principles on the basis of the adopted approaches with respect to:

- the *sources* on which the model evaluates the trustee's trustworthiness,
- the kind of *metric* for trust measures.

Different Kinds of Sources

With respect to the trust sources, as shown in Chapter 6, we can distinguish among *direct experience* and *indirect experience*. Direct experience is the more simple and elementary agent's source deriving from its previous direct experiences with other agents and with the world; it strongly depends on the agent's perceptive apparatus: the kind of input it is able to perceive.

Indirect experience sources can be, in their turn, articulated in the so-called *reputation* (others' experience directly communicated or made available (and possibly mediated) by some central or decentral mechanism) and types of *general reasoning and deduction*, like inference over categories, classes of agents, situations (scripts), and so on.

The sources based on the reasoning and their influences on the trust decision are partially shown in Chapter 6 of this book. Reputation has become a very diffused and practiced

study field; in fact, on the basis of reputation a set of automatic models and systems for attributing trust were directly studied and built in the last 15 years. In many cases the reputational approach represents the only criterion for defining the trustworthiness of the interactive agents.

Reputation mechanisms are distinguished in *centralized* and *decentralized* mechanisms.

Centralized Reputation Mechanisms

Centralized reputation mechanisms are widespread in electronic commerce: *eBay* [eBay site] (Resnick and Zeckhauser, 2002), *Amazon* (Amazon site], and many others' e-commerce systems manage these kinds of mechanisms in which all the users have their own reputational profile stored in a centralized database. In these systems each user, after an interaction (transaction) with other users, reports on the behavior of the other providing appropriate ratings and giving textual comments. These ratings and comments are public and each user can read them before starting a new interaction/business with a specific agent. In the eBay system the rate scale is from $-1, 0, +1$ (respectively negative, neutral and positive). All the ratings are stored centrally and the global reputation value is the sum of all the ratings in the last six months. The main limit of this approach is given by the extreme simplicity of the model. In fact, just one dimension of the trustworthiness is taken in consideration (that is a more complex entity) and the acritical aggregation of the performances do not give account of the possibility of cheating in few interactions maintaining a good reputation value.

To overcome the limits of the reputation systems shown above, *SPORAS* (Zacharia and Maes, 2000) has been developed and it introduces new methods for aggregating the ratings. In particular, the updating of the ratings follows these principles:

1. New users start with a minimum reputation value and they build up reputation during their activity on the system.
2. The reputation value of a user never falls below the reputation of a new user;
3. After each transaction, the reputation values of the involved users are updated according to the feedback provided by the other parties, which reflect their trustworthiness in the latest transaction;
4. If two users happen to interact more than once the system keeps the most recently submitted rating.
5. Users with a very high reputation value experience much smaller rating changes after each update.
6. Ratings must be discounted over time so that the most recent ratings have more weight in the evaluation of a user's reputation.

The six above principles define a more interesting dynamics of the reputation model with respect to more static ones like *eBay*, or *Amazon*. In addition, *SPORAS* introduces a measure of the reliability of the users' reputations: This reliability is based on the deviation of rating values. In this way this system introduces an indication of the predictive power of the algorithm. High deviations correspond to high degrees of variation (or to insufficient activation in the transactions).

Decentralized Reputation Mechanisms

The main problem with the centralized reputation mechanisms is that they are not suitable for open Multi-Agent Systems given that the MAS nature is intrinsically not referred to a central authority, but prevails in the distribution of the various attitudes to the agents who are possibly considering the emergence of the global phenomenon.

An interesting mechanism, developed by Yu and Singh (Yu and Singh, 2003), take into consideration witnesses as information sources. In this way the system (called *referral system*) is based on the individual agents' knowledge and help, exploiting the agent's contacts for each agent in the system. In addition, agents cooperate with each other with respect to these referrals building in fact recommendations to contact other agents. Using referrals each agent has to store a list of agents it knows (acquaintances) with their expertise and to contact them in case of necessity. It is also possible that an agent who is unable to give answer to a query gives back referrals pointing to agents it believes to have the desired information.

A mixed centralized/decentralized approach is introduced by (Jurca and Faltings, 2003). They start from two main questions: Why should agents report reputation information? And, why should they report it truthfully? They think about a set of broker agents, buying and aggregating other agents' reports and selling back reputation information to the agents when they need it. The proposed payment scheme guarantees that agents reporting incorrect or false information will gradually lose money; on the contrary, the honest agents will not lose money.

A reputation model in which the trust evaluation process is completely decentralized has been introduced by (Sabater and Sierra, 2001) and its name is *Regret* (Sabater, 2003). In this model each agent can evaluate others' performance after each direct interaction has recorded its ratings in a local database. In fact in *Regret* the agents can evaluate trust by themselves, without any reference to a central mechanism. The Regret's authors called *direct trust* the trust value derived from these ratings: They are calculated as the weighted means of all ratings (where the weight depends from the recency). Similarly to *SPORAS*, *Regret* measures the predictive power of each trust value through the two reliability measures: the number of ratings taken into account in producing the value of trust and the deviation of these ratings.

In *Regret* the agents can also share their opinions. To this end the system develops a very interesting and sophisticated witness reputation component (for aggregating witness reports). This component strictly depends on the social network built up by each agent: in fact *Regret* uses social network to find witnesses, to select the witnesses for consulation and for weighting the witnesses' opinions. Finally, *Regret* introduces the concepts of *neighbourhood reputation* and *system reputation*. With *neighbourhood reputation* is meant the reputation of the target's neighbour agents: It is calculated by fuzzy rules. With *system reputation* is meant a mechanism to assign default trust values to the target agent based on its social role in an interaction (e.g. service provider, consumer).

Concluding *Regret* uses various sources of trust information, is decentralized, and as a consequence, satisfies the requirements for modelling trust in multi-agent systems. Its main limit is about the fact that it does not specify how the social networks are to be built.

Another interesting reputation model is *FIRE* (Huynh *et al.*, 2006). It is a decentralized model designed for general applications based on a variety of trust sources. In particular, these sources include: *direct experiences* (coming from the agent's interactions), *witness reports* (coming from third-party references), *role-based rules* (provided by end users encoding beliefs about the environment), and *certified reports* (references obtained by the target agent).

The addition of the fourth source (certified reports) to the traditional first three allows trust to measure in situations in which without this source there is no reference.

The *FIRE*'s authors show a set of experimental results in which the performance of their system is really good and better than other developed systems. In particular they show that:

- Agents using the trust measure provided by *FIRE* are able to select reliable partners for interactions and, thus, obtain better utility gain compared to those using no trust measure.
- Each component of *FIRE* plays an important role in its operation and significantly contributes to its overall performance.
- *FIRE* is able to cope well with the various types of changes in an open MAS and can maintain its properties despite the dynamism possible in an environment.
- Although decentralized, to suit the requirements of a trust model in open MAS, *FIRE* still outperforms or maintains a comparable performance level with *SPORAS*, a centralized trust model.

The main problem and limit with *FIRE* is about its assumption that agents are honest in exchanging information with one another: The authors are aware of this limit and are working to introduce reliability measures for witness ratings and certified ratings.

Different Kinds of Metrics

The computational approaches to trust can be classified also on the basis of the different forms of metrics to rate the trust performances. An analysis about the metrics of the different approaches is particularly relevant because it should give account about the underlying assumptions of these approaches.

Discrete or continuous measures (in general in the numerical range $(-1,+1)$ where -1 means no trust and $+1$ full trust) are the most common. But there are also *discrete verbal statements* like those used in (Abdul-Rahman and Hailes, 2000): *Very Trustworthy*, *Trustworthy*, *Untrustworthy*, *Very Untrustworthy*, that give a more direct representation of the human evaluations: The problem is then to translate these statements in adequate misures for the computational process.

We can cite (Schillo *et al.*, 2000) and (Banerrjee *et al.*, 2000) as examples of the use of bi-stable trust values (good or bad) (Ramchurn *et al.*, 2004); while (Witkowsky *et al.*, 2001) proposed the calculation of trust (a continuous measure) that deal with measurable quantities of bandwidth allocation and bandwidth use (they presented a scenario for telecommunications Intelligent Network in which bandwidth is traded by different agents).

Probabilistic approaches have the advantage of exploiting the consolidated apparatus of the probabilistic methods: in particular they can profit from the different derivation methods (from the probability calculus to the advanced statistical methods) (Josang, 2001), (Krukow and Nielsen, 2006).

An interesting and useful approach is given by the Belief Theory that responds to the limits of the probabilistic approaches with respect to uncertain information. In fact, the aim of *belief theory* is to give a formal representation of the *inaccurate* and *uncertain* aspect of information. In this case the sum of the probabilities over all possible outcomes do not necessarily sum up to 1 (the remaining probability can be interpreted as uncertainty).

We have to say that each of these metrics can present both advantages and problems, but what is important in our view is the fact that trust has an intrinsic multi-factorial nature and this peculiar feature has to be represented and implemented also in the quantitative aspects of this phenomenon (see Chapter 3).

One of the more interesting attempts in this direction is represented by the *REGRET* approach where to overcome the mono-dimensionality of the trust performance some fuzziness over the notion of performance itself is used. *REGRET* introduces a rich semantics for the ratings (called *impressions*) by defining their specific features (for example: *delivery date*, *price*, and so on). On the basis of the fuzzy reasoning techniques, the system is able to compose the different dimensions producing a general and global impression of one agent on another. With respect to the fuzzy approach applied to the trust evaluation we showed in Chapter 11 a specific implementation based on the socio-cognitive approach to trust.

Other Models and Approaches to Trust in the Computational Framework

Recent works on designing models for propagating trust in social networks and for selecting the most appropriate services in the service-oriented environments are of particular interest to analyze. In (Yu and Singh, 2003) the models of trust propagation are distinguished by:

- *Trust in expertise*: ability of providing services; and
- *Trust in sociability*: ability of providing referrals.

These two functions are quite different even if relevant: the former is becoming more important given the increasing use of the Internet by people (Internet users are 'now', 40 years after its birth, 1,663 million) and its circulation of both traditional and innovative services.

With regard to the latter, if we use trust to evaluate people and information we have to compute trust between people who do not know one another and expect to achieve the result that each agent in the network will evaluate the trustworthiness of its potential, often anonymous, partners (Golbeck, 2009).

In *CertProp* model (Hang *et al.*, 2009) i a trust propagation model based on three operators (*aggregation*, *concatenation* and *selection*) s introduced to efficiently and accurately propagate trust in social networks. In this approach a social network (system of interacting agents) is modeled as a directed graph with weights on the links: a sort of *social graph*, in which the nodes represent the users and the edges represent the relationships between these users (the social graph representation is quite widespread and used, see also (Ziegler, 2009), (Levien, 2009)). The problem of propagating trust in this network using the introduced operators is interesting even if not all the problems connected with the trust propagation are solved: the already presented question about the so-called 'trust transitivity' (see Chapter 6) remains an issue that has not been well addressed.

In any case this work, defining a set of algebric properties for the three operators, determines trust propagation in an efficient and accurate way (even if in a simplified domain). Another interesting question analyzed in this work is the classical mapping between *opinions* and *evidences*: in the networks the weights are subjective opinions and not objective evidence. Then, the authors propose approaches for transforming opinions in evidence: in particular they motivate a new way of this transformation based on Weber_fechner law (that describes

the relationship between physical magnitudes of stimuli and the perceived intensity of the stimuli). This transformation also allows the idea that the average opinion yields the lower certainty of transformed trust. It helps to reduce the subjectivity in opinion-based datasets so that the evidence-based approaches like *Cert-Prop* can apply.

The attention to the mathematical properties of the operators, fails sometimes to catch the deeper nature of the trust phenomenon. For example, when I receive different, diverging evaluations (about *Y*) from two or more agents (say *J* and *Z*), not only can I discount the degree of trust in *Y* on the basis of *J* or *Z*'s trustworthiness and believability in me (Hang *et al.*, 2009), not only do I have to combine those converging or diverging values with some mathematical 'aggregation', but I have to choose *among different heuristics*, strategies. Not necessarily is the final value of trust a mix of the various values. I may, for example, be a very suspicious and prudent guy (or adopt a prudent strategy), and, although *J* and *Z* say that *Y* is sure and good, since *W* says that *Y* is not good (or sure) I adopt *W*'s view, and put aside *J* and *Z*'s evaluations. Or I might have an optimistic attitude and adopt always the best, more favorable estimation. Or I might have a strong esteem of *Z* (as evaluator) and trust him very much; although *J* and *W* have different opinions I do not care about them, I adopt *Z*'s opinion (trust) without discounting it by combining it with the other evaluations. In sum, there are different possible heuristics in combining (or not) various evaluations; there is not a unique ('rational') equation.

In (Richardson *et al.*, 2003) the trust propagation model allows each user to maintain trust in a small number of other users. This method first enumerates all paths between the user and every other user who has a local belief in a given statement. Then, the belief associated with each path (concatenation operator) is calculated, and combined with the beliefs associated with all paths (aggregation operator). The aggregation operator is the same as the *Cert-Prop's* one while the concatenation operator is different.

Trust metrics[1] compute quantitative estimates of how much trust an agent *X* should have in *Y*, taking into account trust ratings from other agents on the network.

Two main important applications of trust metrics are: *Advogate* (Levien, 2009) and *Appleseed* (Ziegler, 2009). Both these metrics can be classified as *local group trust metrics*. *Local* is intended versus *Global*: where *Global* take into account all peers in the network and the links connecting them; while *Local* trust metrics take into account personal bias. They operate on partial trust graph information (the web of trust for an agent *X* is the set of relationships emanating from *X* and passing through nodes *X* (directly or indirectly) trusts.

Advogate computes a set of accepted nodes in three steps. First, it is assigned a capacity to every node as a function of the shortest path distance from the seed to that node. Second, there is a transformation of the graph, adding extra edges from each node to a special node (called supersink). Third, it is computs the maximum network flow for the new graph: the accepted nodes are those that have a flow across the special node (supersink).

In contrast to *Advogate*, *Appleseed* uses spreading activation (Quilian, 1968). It spreads energy across the graph, and when propagated through a node, divides energy among sucessors based on the edge weights. The main idea in *Appleseed* is to simulate the spectral decomposition and it requires several iterations to converge towards the set of acceptable nodes.

[1] Here considered with a different meaning with respect to the previous section 'Different kinds of metrics in this chapter.

Another interesting work about trust propagation in social networks is focused on a different aspect of trust inference: in particular, the change in trust values in the network and the impact of that change on the results of the existing algorithms. In their work (Golbeck and Kuter, 2009), Golbeck and Kuter show an experimental study to understand the behavior of different trust inference algorithms with respect to the changes occuring in the social network. Their contribution is on different relevant items that they define with the following questions:

'How far does a single change propagate through the network? How large is the impact of that change? How does this relate to the type of inference algorithm?' ((Golbeck and Kuter, 2009), p. 170). They show the relevance of the chosen algorithm in all three questions.

As we have claimed above the problem of selecting trustworthy services on the web is becoming really relevant given the variety and quantity of the offer. In particular, two main problems should be taken in consideration:

a) The trust evaluation of a service should take into consideration the fact that very often a service is a *composed* service (with different providers, functions and responsibilities).
b) The *dynamism* of service providers and consumers (both the needs of the consumers and the providers' quality of service, continuously change).

Singh (Singh, 2003), has coined the term '*service-oriented computing*' for meaning 'the emergence of independent services that can be put together dynamically at run time and possibly across administrative domains' (p.39). In fact, while works exist about modeling trustworthiness of individual service providers, very few results were reached in modeling groups of providers working to a specific composed service. This approach is also important because – like in our model – consider the 'quality' or the 'competence' intrinsic part of trustwortiness and of trust; differently from many models (in game theory, in philosophy, etc.) that want to eliminate this component/dimension of trust to reduce it only to honesty, reliability, morality, etc.

Hafizoglu and Yolum (Hafizoglu and Yolum, 2009), propose a group trust model to understand the behavior of such teams that realize a composed service. Their work is in turn based on the service graphs model (Yolum and Singh, 2004) where graphs are helpful for reasoning about services that are related to each other. In (Hafizoglu and Yolum, 2009), the authors have to cope with the problem that 'the behavior of an agent in teamwork environment may differ from its behavior in single service environment'. In fact, collaboration may have some influence on the agents' performances. So the individual features of the agents in providing specific services are not so useful for selecting the right (best) team of the composed service. The authors analyze a set of possible tendencies for the agents (*ideal behavior*, *group antipathy*, *group motivation*, *colleague effect*, *task effect*, *familiarity effect*) that have influence on the agents' collaborative performances.

Another interesting work on this problem is (Hang and Singh, 2009) in which is proposed a trust-aware service selection model based on a Bayessan network. The model evaluates the service trustworthiness on both direct and indirect (from referrals) experience. The method models causal relationships between services with Bayesan networks.

The main characteristic of this model is that it can deal with incomplete observation (in such a way taking into account the possibility that underlying services may not be exposed to the consumer) for this introduces a specific parameter representing the percentage of missing data.

12.3 Concluding Remarks

It is now clear – even in engineering – that when we build a new technology, for direct human use, in fact we are building a new 'socio-tecnical system' and even a new 'cognition' dealing with and incorporating that mental, pragmatic, and social 'extension'. This is true with mechanical engineering (factories, cars, tractors, etc,) but is much more important with cognitive and social technologies: like computers, web, and their mediation and support of the entire human individual and social activity, from study and learning to work, from friendship and communities to political participation, from market and business to smart learning environments.

We have to design hand in hand with technology the cognitive, interactive, and collective dimensions. More precisely, we have to design technology with those incorporated dimensions. But in order to do this one should have the appropriate understanding of those dimensions and some theoretical abstraction of them and some possible modeling of them. Otherwise we proceed just in an empirical, haphazard (trials and errors) way.

This is why we believe that a deep and complete model of trust (including the cognitive, emotional, decisional, social, institutional) dimension be not just useful but necessary.

In particular, we believe that to support this kind of human computer, human ambient, human robot interaction, and computer-mediated/supported interaction, organization, work, etc. a technology able to deal with typical human cognitive and social features and phenomena (like expectation, intentions, preferences, like emotions, trust, etc. like norms, roles, like institutions, collectives, etc.) must be designed. A technology endowed with autonomous learning, decentralization, acquisition of local and timely information; able to reason and solve problems; endowed with some proactivity and a real collaborative (not just executive) attitude. We think that autonomous computational 'agents' will play a significant role. But if this is true this will make even more central the role of trust and delegation and of their modelling.

References

Abdul-Rahman, A. and Hailes, S. (2000) Supporting Trust in Virtual Communities. In Proceedings of the Hawaii International Conference on System Sciences, Maui, Hawaii, 4–7 January 2000.

Abrams, M.D. (1995) Trusted system concepts. *Computers and Security*, 14 (1): 45–56.

Amazon site, http://www.amazon.com, world wide web.

Anderson, R. (2001) *Security Engineering: A Guide to Building Dependable Distributed Systems*, John Wiley & Sons Ltd.

Antoniou, G. and van Harmelen, F. (2008) *A Semantic Web Primer*, 2nd edition. The MIT Press.

Baecker, R.M. and Buxton, W.A.S. (eds.) (1987) Readings in Human-Computer Interaction. A multidisciplinary approach. Los Altos, CA: Morgan-Kaufmann Publishers.

Barber, S., and Kim, J. (2000) Belief Revision Process based on trust: agents evaluating reputation of information sources, *Autonomous Agents 2000 Workshop on 'Deception, Fraud and Trust in Agent Societies'*, Barcelona, Spain, June 4, pp. 15–26.

Bishop, M. (2005) *Introduction to Computer Security*. Reading, MA: Addison-Wesley.

Blaze, M., Feigenbaum, J., Lacy, J. (1996) Decentralized trust management. In Proceedings of the 1996 IEEE Conferente on Security and Privacy, Oakland, CA.

Card, S.K., Moran, T.P. and Newell, A. (1983) The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Castelfranchi, C. (1996) Reasons: belief support and goal dynamics. *Mathware & Soft Computing*, 3: 233–247.

Castelfranchi, C., de Rosis, F., Falcone, R., Pizzutilo, S. (1998) Personality traits and social attitudes in Multi-Agent Cooperation, *Applied Artificial Intelligence Journal*, special issue on 'Socially Intelligent Agents', 12 (7/8): 649–676.

Castelfranchi, C., Falcone, R., Lorini, E. (2009) A non-reductionist approach to trust. In Golbeck, J. (ed.) *Computing with Social Trust. Human Computer Interaction Series*, Springer.

Castelfranchi, C. and Falcone, R. (1998) Towards a theory of delegation for agent-based systems, *Robotics and Autonomous Systems*, Special issue on Multi-Agent Rationality, Elsevier Editor, 24 (3-4): 141–157.

Castelfranchi, C. and Falcone, R. (1998) Principles of trust for MAS: cognitive anatomy, social importance, and quantification, *Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98)*, Paris, July, pp. 72–79.

Castelfranchi, C. and Tan, Y.H. (eds.) (1999) *Trust and Deception in Virtual Societies*. Kluwer, Dordrecht.

Chellas, B.F. (1990) *Modal Logic: an introduction*. Cambridge University Press. Cambridge.

Cofta, P. (2007) *Trust Complexity and Control*. John Wiley & Sons Ltd.

Cohen, P.R. and Levesque, H.J. (1990) Intention is choice with commitment. *Artificial Intelligence*, 42, 213–261.

Davies, J. (2006) *Semantic Web Technologies: trends and research in ontology-based systems*. John Wiley & Sons Ltd.

Demolombe, R. (1999) To trust information sources: A proposal for a modal logic framework. In Castelfranchi, C., Tan, Y.H. (eds.) *Trust and Deception in Virtual Societies*. Kluwer, Dordrecht.

Dix, A. Flay, J., Abowd, G. and Beale, R. (2004) *Human-Computer Interaction*. Third edition. Prentice Hall.

Dragoni, A. F. (1992) A model for belief revision in a multi-agent environment. In *Decentralized AI - 3*, Y. Demazeau, E. Werner (eds.), 215–231. Amsterdam: Elsevier.

Dubois, D. and Prade, H. (1980) *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, Orlando, FL.

Dubois, D. and Prade, H. (1980) *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, Orlando, FL.

eBay site, http://www.ebay.com, world wide web.

Ellis, J. and Speed, T. (2001) *The Internet Security Guidebook*, Academic Press.

Fagin, R. and Halpern, Y. (1994) Reasoning about Knowledge and probability. *Journal of the Association for Computing Machinery*, 41 (2): 340–367.

Falcone, R., Singh, M., Tan, Y.H. (eds.) (2001) Trust in Cyber-societies. Lecture Notes on Artificial Intelligence, n°2246, Springer.

Falcone, R., Barber, S., Korba, L., Singh, M. (eds.) (2003) Trust Reputation, and Security: Theories and Practice. Lecture Notes on Artificial Intelligence, n°2631, Springer.

Falcone, R., Pezzulo, G., Castelfranchi, C. (2003) A fuzzy approach to a belief-based trust computation. *Lecture Notes on Artificial Intelligence*, 2631, 73–86.

Falcone, R., Barber, S., Sabater-Mir, J., Singh, M. (eds.) (2005) Trusting Agents for Trusting Electronic Societies. Lecture Notes on Artificial Intelligence, n°3577, Springer.

Falcone, R., Barber, S., Sabater-Mir, J., Singh, M. (eds.) (2008) Trust in Agent Societies. Lecture Notes on Artificial Intelligence, n°5396, Springer.

Falcone, R. and Castelfranchi, C. (2001) Social trust: a cognitive approach, in C. Castelfranchi and Y. Tan (eds.), *Trust and Deception in Virtual Societies*, Kluwer Academic Publishers, pp. 55–90.

Foster, I. and Kesselman, C. (2003) *The Grid 2: Blueprint for a new computing*. Morgan Kaufmann Publishers Inc. San Francisco, CA.

Golbeck, J. (ed.) (2009) Computing with Social Trust. Human Computer Interaction Series, Springer.

Golbeck, J. and Kuter, U. (2009) The ripple effect: change in trust and its impact over a social network. In Golbeck, J., (Ed.), *Computing with Social Trust. Human Computer Interaction Series*, Springer.

Grandison, T., and Sloman, M. (2000) A survey of trust in internet application. IEEE Communication Surveys & Tutorials, 3 (4).

Hang, C.W., Wang, Y., Singh, M.P. (2009) Operators for Propagating Trust and their Evaluation in Socila Networks. In Proceedings of the 8th International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS).

Hewitt, C. and Inman, J. DAI betwixt and between: from 'intelligent agents' to open systems science, *IEEE Transactions on Systems, Man, and Cybernetics*. Nov./Dec. 1991.

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

Jones, A.J.I., Firozabadi, B.S. (2001) On the characterization of a trusting agent: Aspects of a formal approach. In Castelfranchi, C., Tan, Y.H. (eds.) *Trust and Deception in Virtual Societies*, pp. 55–90. Kluwer, Dordrecht.

Jonker, C., and Treur, J. (1999) Formal Analysis of Models for the Dynamics of Trust based on Experiences, Autonomous Agents '99 Workshop on 'Deception, Fraud and Trust in Agent Societies', Seattle, USA, May 1, pp. 81–94.

Josang, A. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 9 (3): 279–311, 2001.

Josang, A. (2007) Trust and Reputation Systems, in Aldini, A., Gorrieri, R. (eds.), Foundations of Security Analysis and Design IV, FOSAD 2006/2007 Tutorial Lectures. Springer LNCS 4677.

Jøsang, A., and Ismail, R. (2002) *The Beta Reputation System*. In the proceedings of the 15th Bled Conference on Electronic Commerce, Bled, Slovenia, 17-19 June 2002.

Jurca, R. and Faltings, B. (2003) Towards incentive-compatible reputation management, in Falcone, R. et al. (eds.), *Trust, Reputation and Security: Theories and Practice*, LNAI Vol. 2631 (pp. 138–147), Sprinter-Verlag.

Kosko, B. (1986) Fuzzy cognitive maps. *International Journal Man-Machine Studies*, 24: 65–75.

Krukow, K. and Nielsen, M. (2006) From Simulations to Theorems: A position paper on research in the field of computational trust. In Proceedings of the Workshop of Formal Aspects of Security and Trust (FAST 2006), Ontario, Canada, August 2006.

Lesser, V.R. An overview of DAI: viewing distributed AI as distributed search. *Journal of Japanese Society for Artificial Intelligence.* Special issue on Distributed Artificial Intelligence, 5 (4): 392–400, 1990.

Levien, R. (2009) Attack-resistant trust metrics. In Golbeck, J., (Ed.), *Computing with Social Trust. Human Computer Interaction Series*, Springer.

Liau, C.J. (2003) Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. *Artificial Intelligence*, 149: 31–60.

Lorini, E. and Demolombe, R. (2008) From binary trust to graded trust in information sources: A logical perspective, pp. 205–225. In Falcone, R., Barber, S., Sabater-(Mir and Singh, 2008) Mir, J. and Singh, M. (eds.) (2008) Trust in Agent Societies. Lecture Notes on Artificial Intelligence, n°5396, Springer.

Marsh, S. P. (1994) Formalising Trust as a computational concept. PhD thesis, University of Stirling. Available at: http://www.nr.no/abie/papers/TR133.pdf.

Motro, A. and Smets, Ph. (1997) *Uncertainty Management in Information Systems: from needs to solutions*. Klewer, Boston.

Pezzulo, G. and Calvi, G. (2004) AKIRA: a Framework for MABS. *Proc. of MAMABS 2004.*

Quilian, R. (1968) Semantic memory, In Minsky, M., (editor), *Semantic Information Processing*, MIT Press, Boston, MA, USA, pp. 227–270.

Ramchurn, S., Huynh, T. D., Jennings, N. R. (2004) Trust in multi-agent systems. *The Knowledge Engineering Review*, Cambridge University Press, 19 (1): 1–25.

Rasmussen, L. and Janssen, S. (1996) Simulated Social Control for Secure Internet Commerce. In Meadows, C. (Editor), Proceedings of the 1996 New Security Paradigm Workshop. ACM.

Resnick P. and Zeckhauser, R. (2002) Trust among strangers in internet transactions: Empirical analysis of eBay's reputation systems. In Baye, R. (Editor), The economics of the internet and e-commerce. Vol. 11 of *Advances in Applied Microneconomics*. Elsevier Science.

Richardson, M., Agrawal, R., and Domingos, P. (2003) Trust Management for the Sematic Web, in the Second ISWC, LNCS 2870, pp. 351–368.

Sabater, J. (2003) Trust and Reputation for Agent Societies, PhD thesis, Universitat Autonoma de Barcelona.

Sabater, J. and Sierra, C. (2001) Regret: a reputation model for gregarious societies, In 4th Workshop on Deception and Fraud in Agent Societies, (pp. 61–70). Montreal, Canada.

Schillo, M., Funk, P., and Rovatsos M. (1999) Who can you trust: Dealing with deception, Autonomous Agents '99 Workshop on 'Deception, Fraud and Trust in Agent Societies', Seattle, USA, May 1.

Schneier, B. (1996) *Applied Cryptography*, 2nd edn, Wiley.

Shafter, G. (1976) *A Mathematical Theory of Evidence*. Princeton University Press: New Jersey.

Shoham, Y. and Leyton-Brown, K. (2008) *Multi-agent Systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Singh, M.P. (2003) Trustworthy service composition: challenger and research questions. In Falcone, R., Barber, S., Korba, L., Singh, M., (Eds.), Trust Reputation, and Security: Theories and Practice. Lecture Notes on Artificial Intelligence, n°2631, Springer.

Skevington, P. and Hart, T. Trusted third parties in electronic commerce, *BT Technology Journal*, 15 (2), 1997.

Skogsrud, H., Benatallah, B., and Casati, F. (2003) Model-driven trust negotiation for web services. *IEEE Internet Computing*, 7 (6): 45–52.

Smith. R.G. (1980) The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*.

Stallings, W. (1999) *Cryptography and Network Security*, Prentice Hall.

Stallings, W. (2001) *SNMP, SNMPv2, SNMPv3, and RMON 1 and 2*, 3rd edn, Addison Wesley.

Uszok, A., Bradshow, J. M., and Jeffers, R. (2004) KAoS: A policy and domain services framework for grid computing and semantic web services. In Jensen, C., Poslad, S., and Dimitrakos, T. (eds.), *Trust Management: Second International conference*, iTrust 2004, Oxford, UK, March 29–April 1, 2004. Lecture Notes in Computer Sciente n°2995, (pp. 16-26), Sprinter-Verlag, Berlin, Heidelberg.

Weiss, G. Distributed artificial intelligence meets machine learning: learning in multi-agent environments, Berlin: Springer-Verlag, Lecture notes in *Computer Science*, vol. 1237, 1997.

Wong, H.C., Sycara, K. Adding security and trust to multi-agent systems, Proceedings of Autonomous Agents '99. (Workshop on Deception, Fraud and Trust in Agent Societies), 2000.

Wooldridge, M. (2002) *An Introduction to Multi-agent Systems*, John Wiley & Sons Ltd.

Yu, B. and Singh, M.P. (2003) Searching social networks. In Proceedings of the second international joint conference on autonomous agents and multi-agent systems (AAMAS). Pp. 65-72. ACM Press.

Zacharia, G. and Maes, P. Trust management through reputation mechanisms. *Applied Artificial Intelligence Journal*, 14 (9): 881–908, October 2000.

Ziegler, C.N. (2009) On propagation interpersonal trust in social network. In