Defining Trust and E-Trust

Article · August 2011

DOI: 10.4018/jthi.2009040102

Mariarosaria Taddeo
University of Oxford

DEFINING TRUST AND E-TRUST: FROM OLD THEORIES TO NEW PROBLEMS

Mariarosaria Taddeo; Faculty of Philosophy, University of Padua, Italy; IEG, University of Oxford; GPI, University of Hertfordshire; Fellow of the International Society of Ethics and Information Technology (INSEIT)

mariarosaria.taddeo@philosophyofinformation.net

ABSTRACT

The paper provides a selective analysis of the main theories of trust and e-trust (that is, trust in digital environments) provided in the last twenty years, with the goal of preparing the ground for a new philosophical approach to solve the problems facing them. It is divided into two parts. The first part is functional toward the analysis of e-trust: it focuses on trust and its definition and foundation and describes the general background on which the analysis of e-trust rests. The second part focuses on e-trust, its foundation and ethical implications. The paper ends by synthesising the analysis of the two parts.

INTRODUCTION

Although trust is largely recognised as an important issue in many fields of research, we still lack a satisfactory definition and foundation of it. Moreover, in recent years, the emergence of trust in digital contexts – known as *e-trust* – has created new theoretical problems. So the first part of this paper, dealing with the old problems of trust, is functional toward the analysis of e-trust and it is meant to describe the general background on which the analysis of e-trust rests. The second part focuses on the new problems posed by e-trust. Their investigation is meant to prepare the ground for a new philosophical approach that might overcome the highlighted difficulties. Let me now provide a more detailed summary.

I first briefly describe Luhmann's contribution to the analysis of trust and then focus on the definition of trust provided by Gambetta. This definition has provided a general understanding of trust and influenced much of the literature. As we shall see, it stresses two aspects: the decision that an agent takes to trust and the relation between trust and risk. I will argue that, despite several valuable features, the definition still faces two main problems: the specification of the parameters that determine the decision to trust, and the explanation of the reasons why an agent should decide to take the risk of trusting another agent. These are problems that affect the analysis both of trust and of e-trust.

In the second part of the paper, I discuss the problems that affect e-trust more specifically. The first problem is whether e-trust is possible at all. Some of the literature has denied that trust in digital environments may ever occur. This position rests on the assumption that "trust needs touch" – that it needs to be based on direct physical interaction, which of course does not exist in digital contexts. In the next two sections, I draw attention to two other problems: the role of e-trust in the dynamics of a distributed artificial system, and the relation between e-trust and the ethical values that AAs might be endowed with.

In the last section, I conclude by pulling together the different threads of the analysis in order to summarise the problems left unsolved.

TRUST: A DECISION MAKING PROCESS

Trust is often understood as a relation between an agent (the *trustor*) and another agent or object (the *trustee*). The relation is supposed to be grounded on the trustor's beliefs about the trustee's capabilities and about the context in which the relation occurs. This is a generalisation of the definition of trust provided by (Gambetta, 1998). Before exploring in more depth Gambetta's analysis, however, let me briefly recall some of the more relevant points in Luhmnann's analysis of trust, (Luhmann, 1979). This analysis should be considered the starting point for the modern approach to trust and its cognate concepts.

Luhmann examines the function of trust and the social mechanisms through which trust is generated. He specifies the reason why society in general needs trust. Such a need rests on the fact that trust is a starting point for the derivation of rules for proper conduct, or for ways of acting successfully by reducing complexity and uncertainty in a given social system. Following Luhmann "trust is an effective form of complexity reduction", (p. 8).

For Luhmann, trust is a decision taken by the trustor on the basis of the following parameters: familiarity, expectation and risk. Familiarity is the acquaintance of the trustor with the potential trustee and with the systems. It is the variable that provides a reliable background for the trustor's choice to trust. Expectation is the reason for which an agent decides to trust. For Luhmann, trust is present only when the expectation to trust makes a difference to a decision, otherwise what one has a simple hope. Trust is a risky investment. Following Luhmann, this is so because to trust is to take a decision and risks are a component of decision and action.

This brief overview of Luhmann's analysis underlines the main issues present in any attempt to investigate trust: (a) trust as a result of a decision process, (b) the need of a reliable background as a necessary requirement to trust, (c) the expectation and (d) the risk related to the choice to trust. All these issues have been addressed in the theories analysed in the rest of this paper; particular attention to trust as a result of a decision process has been paid by Gambetta's analysis.

Gambetta defines trust as follow: "trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action", (p. 216).

According to Gambetta, trust is grounded on the probabilities attributed by the trustor to her own beliefs about the trustee's behaviour and abilities. Once calculated, the probability is compared with a threshold value and placed on a probabilistic distribution, where complete distrust is equal to 0, complete trust is equal to 1, and the mid-point (0.50) represents uncertainty. It is only if the level of probability is equal to, or higher than, the threshold value, that a trustor *a* starts to trust a trustee *b*. This is because the probability is also a measure of the risk for *a* that *b* will not act as *a* believes (or perhaps merely hopes) *b* will. The higher the belief's probability, the lower the risk of the trustee's misbehaviour. A high probability of belief guarantees a low risk for the trustor.

At first sight, Gambetta's definition may seem satisfactory. A deeper analysis, however, uncovers its limitations. Gambetta's definition focuses only on the decision process behind trust, missing other relevant aspects, like how trust affects the behaviour of the agents involved and the effects of their performances. When an agent *a* trusts, *a* not only holds some beliefs about the trustee *b*, but also *behaves* in a specific way towards *b*. Most importantly, if *a* relies on *b* to perform some action, then *a* might, because of her trust, not supervise *b*'s performance of that action. Thus, trust can benefit the trustor by allowing her to save resources in achieving a goal, or harm her if the trustee fails to behave as expected, impeding the trustor in achieving that goal. This deficiency of Gambetta's definition limits its explanatory power. Gambetta is right in emphasising the correlation between trust and risk, and between risk and the probabilities attached to the trustor's beliefs. However, without considering the possible advantages of trusting, it remains unclear why one agent would decide to trust another and accept the risk of doing so. So, it appears that Gambetta's definition provides a starting point in the analysis of trust, but that it is not a fully satisfactory analysis.

One might object that, for Gambetta, trust is directed towards cooperation, so that the aspects of trust unaccounted for, in the definition, might be explained by referring to the proprieties of cooperative relations. However, trust and cooperation are different kinds of relations, and one cannot explain the features of the former by reducing them to features of the latter. Consider cooperation: cooperative agents distribute the tasks among themselves to reach, in the best possible way, a shared goal. This presupposes that the agents are aware of their roles in cooperation. The fact that something or someone *b* is part of a trust relation, however, does not entail that *b* is aware of it. A trustee *b* may not know that she is being trusted. This becomes trivially true when a trust relation includes a human trustor and an object as a trustee. I may trust the elevator, but we are clearly not collaborating. Indeed, trust and cooperation do not overlap and may easily give rise to different kinds of interactions. By referring to cooperative behaviour in order to explain the peculiarities of trust relations, Gambetta's analysis fails to identify the peculiarities of the latter.

The relation between trust and cooperation, described by Gambetta, leads to a second objection. According to Gambetta, whenever *a* trusts *b*, *a* is actually attempting to cooperate with *b*. If the conditions required by mutual trust are fulfilled, cooperation arises. This makes trust a necessary condition for cooperation. Yet this is untenable, for it is clear that there are both cases of mutual trust without cooperation and cases of cooperation not based on trust. Competitive situations are good examples of cases in which participants may trust each other while knowingly and purposefully refusing to cooperate. Consider two politicians from opposite parties: they may trust each other not to commit electoral fraud but, despite this, compete against each other in order to be elected, and hence refuse to cooperate to lead the country together. Likewise, it is easy to imagine examples of cooperation that do not presuppose trust. During World War II, the USA and USSR did not trust each other (e.g. on the management of the political equilibrium in Europe) but still cooperated to fight the Nazi-Fascist alliance. In such cases, common goals may lead to the creation of unusual bedfellows, who cooperate without trusting each other.

To summarise, cooperation might arise from trust, but its doing so is not necessary and depends rather on the context, goals and intentions of the agents involved. Any analysis that grounds trust on cooperation is bound to be unsatisfactory. Further problems arise when one focuses on e-trust.

TRUST IN DIGITAL ENVIRONMENTS

In this and in the following sections I will provide an overview of the debate on e-trust and on the general understanding of this phenomenon. To do so, I will disregard those analyses of e-trust which focus only on its specific occurrences, such as, for example, e-trust in business ethics, in e-commerce, in system management.[1]

E-trust occurs in environments where direct and physical contacts do not take place, where moral and social pressures can be differently perceived, and where interactions are mediated by digital devices.[2] All these differences from the traditional form of trust give rise to a major problem that any theory of e-trust must solve: whether trust is affected by environmental features in such a way that it can only occur in non-digital environments, or it is mainly affected by features of the agents and their abilities, so that trust is viable even in digital contexts. There are different positions taken in the literature about this problem. Let me start by describing three classical arguments that connect trust to features of the environment and so deny the emergence of trust in digital environments.

All the arguments against the existence of e-trust assume that the emergence of trust requires three conditions that cannot be fulfilled in digital environments. These conditions are:

1. direct interactions between the agents;
2. the presence of shared norms and ethical values that regulate the interactions in the environment;
3. the identification of the parts involved in the interactions.

The detractors of e-trust, for example (Nissenbaum, 2001), consider the features of digital environments an obstacle to the fulfilment of these conditions.

The first obstacle concerns the definition of rules and norms that govern interactions in digital environments. According to the critics of e-trust, trust is only possible within the norms and values that regulate a community life. However, these values are culturally and geographically defined. So it does not seem possible to find them in the interactions of virtual communities, which are often de-localised and may be strongly multicultural.

The second obstacle consists in the identification of the interacting agents. In digital environments, agents can remain anonymous and can often be difficult to identify. This makes deception easier and might diminish the agents' sense of responsibility. In such environments, the risk of deceit might be higher than usual, and it might therefore be much more difficult for agents to trust one another. It seems irrational to trust someone without knowing her identity.

This brings us to the third obstacle, which can be summarised by the expression "trust needs touch". Trust needs, or at least is encouraged by, direct, visual, and physical acquaintance among the agents involved. Since, by definition, these kinds of interactions are absent in digital environments, it is argued that trust cannot occur in such environments.

Despite their *prima facie* plausibility, all of these objections against e-trust can be rebutted. Trust in digital environments is mostly associated with trust over the Internet and especially with that of e-commerce. In these cases, e-trust is often reduced to a matter of security. Consider, for example, the way e-trust is approached in trust management, see (Blaze, Feigenbaum, & Lacy, 1996) and (Jøsang, Keser, & Dimitrakos, 2005). One objection is that, even if personal identification, ethical values and direct contacts help form beliefs about a trustee's intentions, they are not necessary requirements, either for trust in general or for e-trust in particular. An agent can form the beliefs that allow her to trust an agent or object (e.g. an automated service) without interacting directly with the trustee. Consider referral-based trust in real environments. This kind of trust is based only on communication processes. It is the kind of trust that one develops in an unknown agent by considering only the recommendations about that agent provided by other agents or by other information sources, such as newspapers or televisions. Referential trust is one of the main kinds of trust developed in digital environments in which communication processes are easily performed.

The three requirements described above can actually be fulfilled in digital environments and that is why, as a matter of fact, e-trust is becoming increasingly common. Nowadays, many tools allow users to identify the agents they virtually interact with. Consider peer-to-peer programs, mailing lists, chats and blogs and the many features or services going under the name of Web 2.0 applications. The people who take part in these interactions have to provide an email address through which they can be identified. The number of internet communities is growing, too. These communities develop shared norms to regulate the behaviour of their members, despite being spread around the world and containing different cultural perspectives. In short, the three conditions listed above are neither necessary constraints on trust nor unfulfillable requirements online. Unsurprisingly, they are not obstacles to the emergence of e-trust.

The debate about e-trust does, however, identify a major problem with the foundations of e-trust. If e-trust does not rest on direct physical interaction, it must rest on some other grounds, for instance, information or the agent's attitude. So, despite the fact that e-trust is a perfectly reasonable and common phenomenon, a theory of e-trust must explain what these other grounds are.

E-TRUST AS AN ATTITUDE

One way to explain how trust can emerge in digital contexts has been described by Weckert in (Weckert, 2005). His analysis seeks to explain the occurrences of e-trust by grounding them on agents' attitudes to trust.[3]

In quoting Baier,[4] Weckert states that trust is "the cognitive, the affective and the conative. Trust is any of these, is all three", (Weckert, 2005). Weckert considers trust a paradigm through which one acts with another agent and with the environment. Generally speaking, for Weckert, to trust an agent is to see her as if she were trustworthy. In this sense, trust is an agent's attitude. The cognitive aspect, namely the trustor's beliefs, plays only a complementary role in defining a trustee's trustworthiness. What grounds trust in the first place is an attitude of the trustor. This one comes from the natural inclination of the trustor and can be facilitated by her moral values.

Weckert considers the trustor's attitude to be a crucial condition for the emergence and preservation of trust. He claims that, once an agent has a trusting attitude, she will still trust, even after being deceived.

According to Weckert, if trust were a purely cognitive state, defined only by a trustor's beliefs and expectations, then it would be difficult to explain why someone would take the risk of trusting agents whose trustworthiness is not evident. In support of his thesis, he cites the emergence of trust in digital environments. According to him, e-trust is based on the trustor's tendency to see other agents as if they were trustworthy rather than on the rational evaluation of the trustees' behaviour. In this case, the trustor *a* chooses to act as if she trusts, delaying the rational evaluation of beliefs and risks: *a* will evaluate the outcomes of her choice at a second stage. If it turns out that the virtual environment favours positive outcomes, then genuine trust, based on the trustor's beliefs, will emerge.

Weckert's argument highlights some important features of trust online, but faces one difficulty. In acting as if she trusted, the trustor *a* actually *does trust* another agent *b*. This much is acknowledged by Weckert: *a* actually interacts with *b*, and takes what is presumably a higher risk, given the lack of information about *b* and the environment. This is not a problem *per se*, but it leaves unexplained a crucial issue: why any agent should decide to engage in the very dangerous behaviour of unjustified trust. For this reason, considering the emergence of e-trust to be a consequence of an agent's attitude does not provide an explanation of why e-trust emerges in the first place. Empirically, we know that it is often the case that, before trusting, an agent considers (or tries to consider) all possible variables that could affect the outcome of her trusting.

In defence of Weckert's position, one might argue that e-trust emerges in environments in which the decision to trust cannot be based on the probabilities of the beliefs or on the calculation of the outcomes, because the parameters, to which an agent would refer in making these calculations, are not available in digital environments. Given this premise, Weckert's analysis may seem to provide the best explanation of trust in digital contexts, yet this is only partly true. Consider e-commerce. E-commerce is a diffuse phenomenon, which has grown exponentially in recent years and which continues to expand even though the risks that accompany it are high and largely well known. Users of e-commerce trust the websites, the companies, the pay systems, and sometimes other users whom they have never seen or met before. Weckert would explain these occurrences of e-trust by referring to a non-rational choice, made by the pioneers of e-commerce, which has later been reinforced by positive outcomes. However, recent studies[5] contradict Weckert's position, for they show that on-line purchases, especially when made for the first time, are based on carefully determined calculations. There are well-defined parameters – such as the brand, the technology of the web site, the seals of approval, and previous experiences made by other customers – that have to be fulfilled to make users develop a level of e-trust high enough to decide to purchase something on-line. Hence, although Weckert's analysis acknowledges the importance of on-line contexts, it fails to give an explanation of e-trust, and leaves open both the problem of its emergence and the problem of the role that trust can play in on-line interactions.

E-TRUST IN ARTIFICIAL DISTRIBUTED SYSTEMS: AAS

So far, I have considered the occurrence of e-trust in hybrid contexts, in which both human and artificial agents interact. I will now consider a different occurrence of e-trust: that between the AAs of a distributed system.

Although the term AAs is used in different areas of research, there is not a single universally accepted definition. Therefore, before proceeding further in the description of the analyses of e-trust among AAs, I will briefly recall two of the more common definitions of AAs provided in the literature. The goal of this section is not to review the literature on AAs but to provide the reader with some guidelines in the understanding of the use of AAs in the rest of the paper.[6]

There are several factors to take into consideration in defining AAs, the environment, the flexibility, the autonomy and the reactivity of the agent. All these parameters are taken into consideration in the definition provided in (Wooldridge & Jennings, 1995). The authors state that "a hardware or (more usually) software-based computer system that enjoys the following properties:

- autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: agents interact with other agents (and possibly humans) via some kind of agent-communication language;

- reactivity: agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the Internet, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it;
- pro-activeness: agents do not simply act in response to their environment, they are able to exhibit goal-directed behaviour by taking the initiative." (p. 2)

Another definition, which includes all the parameters mentioned above has been provided by (Floridi & Sanders, 2004). The authors define AAs as an entity that is interactive (able to respond to the environmental stimuli); autonomous (able to change its states according to its own transaction rules and in a self-governed way independently from the environmental stimuli), and adaptable (able to change its own rules of transaction according to the environment).

AAs often operate within distributed systems in which the constituent components are spread throughout a network, and are subject to constant change throughout the system's lifetime. Examples include peer-to-peer computing (Oram, 2001), the semantic Web (Berners-Lee, Hendler, & Lassila, 2001), Web services and e-business, autonomous computing and the grid (Foster & Kesselman, 1998). Distributed systems can be modelled as multi-agent systems (MAS) that are composed of autonomous agents that interact with one another using particular mechanisms and protocols. In such a system, e-trust plays a fundamental role, because both the system and the agents may have limited computational and storage capabilities that restrict their control over interactions. Moreover, the limited bandwidth and speed of communication channels limit the agents' sensing capabilities in real-world. Thus, in practical contexts it is usually impossible to reach a state of perfect information about the environment and the interaction partners' properties, possible strategies, and interests, see (Axelrod, 1984) and (Russell & Norvig, 1995). AAs are therefore necessarily faced with significant degrees of uncertainty in making decisions. In such circumstances, agents have to trust each other in order to minimise the uncertainty associated with interactions in open distributed systems.

Having described AAs and the importance of e-trust we can proceed further to the analysis of e-trust among AAs of distributed systems.[7]

E-trust among AAs

An analysis of trust among AAs in multi-agent systems (MAS) has been provided by Castelfranchi e Falcone in (Castelfranchi & Falcone, 1998). Its purpose is to prove, first, that e-trust is a mental state of an AA and, second, that e-trust is the mental background of delegation.

According to the authors "trust is a *mental state*, a complex *attitude* of an agent *x* towards another agent *y* about the behaviour/action *a* relevant for the result (goal) *g*", (Castelfranchi & Falcone, 1998). The "mental state" of e-trust is based on the trustor's beliefs about the trustee's attitudes concerning the relationship between the trustee and the trustor.

Castelfranchi and Falcone take e-trust to be a threshold value that is the result of a function of the subjective certainty of the beliefs held by an AA. Only if the level of trust is higher than a given threshold will the trustor decide to delegate the execution of a given action to the trustee. By assuming this premise, the authors state that e-trust is the "mental counter-part of delegation", (p.74) because the trustor chooses to delegate a given action by relying on her own beliefs. Finally, they claim that if trust occurs between two agents, then the trustee will be "committed to *x* [the trustor] to do *a*." There is "an (explicit or implicit) promise to do so which implies interpersonal duty." (p. 78)

Let us now consider three shortcomings of the analysis provided by Castelfranchi and Falcone.

First of all, the definition of e-trust as a mental state of an AA is incorrect. It is anthropomorphic, for it attributes to AAs features – such as self-confidence, a complex attitude, willingness and the capacity to make promises – that are incompatible with our current and foreseeable technology and the classic definition of AAs, according to which an AA is "a computer system that is situated in some environment and that is capable of autonomous actions in this environment in order to meet the design objectives", (Wooldridge, 2002).[8]

A second objection concerns the relationship between e-trust, trust, and delegation. Like the definition provided by Gambetta, the analysis provided by Castelfranchi and Falcone turns out to be too reductive. It does not account for proprieties related to interactions based on trust, but mistakenly takes them to be forms of delegation. The authors describe e-trust as a measure of the beliefs that *a* holds *when a* intends to delegate something to *b,* another agent or an object. Yet cases in which trust occurs without delegation are perfectly possible and indeed common. Recall the examples of the politicians introduced above: they might trust each other, but neither of them would delegate the task of leading the country to the other. It is true that if someone has to delegate something, then she will prefer to choose someone she trusts. But this does not entail that trust always has delegation as its final or defining goal.

One could respond by observing that an agent trusts when she needs another agent or object to perform a given action, and hence that some form of delegation is the obvious and necessary consequence of trust. This argument is indeed offered by Castelfranchi and Falcone. It assumes that a trustor has a "dependence belief": the trustor *a* needs the trustee *b* and believes that "[she] depends on the trustee' to achieve her goal" (p.75).

This argument, however, is based on the erroneous assumption that the trustor is dependent on an agent or object to perform an action, and that, given this necessity, she will delegate this action to someone whom (or something that) she trusts. By assuming dependence, delegation follows straightforwardly. But it is not true that the trustor always depends on the trustee. Let us consider again the example of the two politicians: the politicians trust each other but they do not depend on each other. Or consider a process in which *a* negotiates with *b* to purchase an item, as this might happen on eBay: *a* trusts *b* to be an honest seller and hence not to sell a damaged item. The trust of *a* in *b* does not entail a dependence of *a* on *b*. Hence, dependence on the trustee is not necessary for trust. But if dependence is not entailed by trust, then neither is delegation.

The last criticism concerns the trustee's promise to satisfy the trustor's needs. The promise is based on two assumptions. First, that *b* is aware that it is being trusted and, second, that trust influences *b*'s behaviour. The first assumption is false. The trustee's awareness is a possible but unnecessary feature of trust: one can be trusted without being aware of it. As for the second assumption, this conflicts with the uncertainty of the outcomes of trust assumed by the definition. If trust can bias the trustee's behaviour, and if it can influence the trustee in order to make it satisfy the trustor's needs, then the outcomes of trust can no longer be uncertain, or must at least be much less so. Because of its trust, *a* is confident that *b* will fulfil its expectations. On the basis of these two assumptions, the trustee's promise to act according to the trustor's expectations is not justified.

From the argument of Castelfranchi and Falcone it follows that the trustee abides by some ethical principles – at least the principle to respect promises. Even if the objection shows that trust does not presuppose the commitment of the trustee, the question arises whether e-trust needs to be based on ethical requirements. The role of moral values on the emergence of trust is an important issue, so I will examine it further in the next section.

NORMATIVE E-TRUST[9]

The theory of e-trust examined in this section has been provided by Tuomela and Hofmann in (Tuomela & Hofmann, 2003). The authors' analysis grounds e-trust in the ethical principles of the AAs of a distributed system. Like Gambetta, they consider e-trust to be necessary for the development of cooperation and the formation of groups, i.e. MAS.

The authors concentrate on rational social normative trust (hereafter *normative trust*), distinguishing it from *predictive trust* and *predictive reliance* between agents. They describe trust as a relationship based on the trustee's trustworthiness but claim that it also requires the trustor's belief about her dependence on the trustee's performances.

Normative trust presupposes existing relationships among the agents based on mutual respect, social rights and moral norms. Given these conditions, the trustor *a* decides to depend on the trustee *b* performing an action, and feels comfortable with this dependence because she believes that *b* is committed to that action by the moral relationship that exists among them. In saying that *a* feels comfortable about depending on *b*, the authors mean that *a* does not feel that she is taking a risk in trusting *b*. Normative trust is distinguished from predictive trust in that the latter is said to be the trustor's expectation that the trustee's actions will be beneficial to the trustor unintentionally, and not owing to any social or moral normative value. In this kind of trust, the trustor relies on the context and only some of the trustee's features.

According to Tuomela and Hofmann, only normative trust deserves the label of trust *tout court*: predictive trust and predictive reliance should be described in terms of the more general notion of reliance.

Let us now look at normative trust in more detail. Tuomela and Hofmann describe a set of necessary and sufficient conditions for the occurrence of normative trust. They are as follows. First, *a* does not intend to perform an action *x* by herself. Second, *a* is interested in *b*'s performing *x*. (In this sense, she depends on *b*'s performances.) Third, *b* can be influenced in her performances by knowing that *a* depends on her.

Tuomela and Hofmann then specify five necessary and sufficient conditions for normative trust, including requirements on the beliefs of the trustor about the trustee's performances, skills, and goodwill. They also include the requirements that the trustor holds "positive feelings" about her dependence on the trustee and that she has the "attitude" to depend on the trustee to achieve a given goal.

The condition about the trustor's belief in the goodwill of the trustee is the normative aspect of normative trust. The trustor expects the trustee's goodwill based on their relationship of mutual respect. Goodwill does not mean that *b* is generally good-willed towards *a*; rather, it means that *b*'s behaviour towards *a* is biased by genuine caring and by moral reasons. In normative trust, the social or moral grounds justify the trustor's expecting good-willed behaviour from the trustee. The authors claim that social and moral duties are "a glue made of strong ingredients", (Tuomela & Hofmann, 2003) which commits the trustee to performing a given action. They identify the social and moral ground with the mutual respect that must occur in the interactions of the agents. Based on mutual respect, the trustor accepts her dependence.

The authors stress that, in the case of normative trust, there is no longer any uncertainty: the outcomes of trust are the consequences of the social and moral norms that govern the pre-existing relationships of the agents.

Two criticisms can be moved to this analysis. By formulating them I will reject Tuomela and Hofmann's claim that normative trust is the only relevant type of trust and argue, instead, that normative trust is only one aspect of a more general notion.

The first criticism turns on the connection between trust and dependence. Like Castelfranchi and Falcone, Tuomela and Hofmann consider the main feature of trust to be some kind of dependence and, more specifically, the attitude with which an agent views her dependence on another. But this explanation is too restrictive, for we have seen that there are many, genuine instances of trust without any dependence.

One may object that the main trait of trust is not dependence, but the way the trustor deals with her interaction with the trustee. What distinguishes trust from the more general idea of reliance is that, owing to social and moral relationships, the trustor is justified in "feeling comfortable" about interacting with the trustee.

The second criticism allows one to overcome this objection. The trustor's decision to trust depends on her beliefs and the context in which the interaction occurs. It is arguable that, whenever an agent decides to trust, she is comfortable with the idea of interacting with the trustee, even if there are no norms that guarantee the trustee's respect.

In conclusion, it is possible to have trust-based interactions even when social and moral norms are not present. Hence, social and moral norms cannot be considered a necessary requirement for the development of trust in relationships. The authors' claim that only normative trust deserves to be called trust remains unsupported.

CONCLUSION: OLD AND NEW PROBLEMS

In the previous sections, I provided an overview of the theoretical literature on trust and e-trust. I showed that, though important aspects of these phenomena have been explained, such as the relation between trust and risk, many other features need to be clarified. I focussed on four problems affecting current theories of trust and e-trust:

1. the definitions of trust and e-trust. Trust and e-trust have been defined in the following ways: (i) as a probabilistic evaluation of trustworthiness, (ii) as a relationship based on ethical norms, and (iii) as an agent's attitude. All of these definitions focus only on the trustor's beliefs, and so give a partial explanation of the

phenomenon of trust. They leave many questions unanswered. They do not clarify what the effects of trust on the involved agents' behaviour are, nor for what reason an agent decides to trust. All the analyses of trust provided agree in considering trust necessary for the development of a social system, but none of them explains the reasons why trust has such an important role;

2. the occurrences of e-trust. Some of the literature questions whether it is possible for trust to develop in digital environments. The belief that it is not rests on the assumption that "trust needs touch" – that it must be based on direct physical interaction. For those who hold this assumption, trust cannot arise merely from the information that the trustor holds about other agents and about context. Yet, this is clearly false, for we witness e-trust daily, when we use the Internet and other the digital devices. Some further explanations is required;
3. the role that e-trust might have in the dynamics of distributed artificial systems. The question is whether e-trust gives any kind of advantage to the AAs that trust;
4. the relation between e-trust and ethical principles. On the one hand, e-trust seems to rest upon some ethical principles, insofar as an agent's loyalty or honesty online begets her trustworthiness. On the other hand, e-trust seems to be more related to an agent's capacities to perform a given task, and trust relations seem to be grounded on practical rather than ethical principles. A theory of trust and e-trust, even if not ethically biased, should explain what the role of ethical and practical values is.

These four problems will need to be solved by any satisfactory theory of trust and e-trust.

ACKNOWLEDGMENTS

I discussed several drafts of this paper in two meetings, the *Twenty fourth North American Conference on Computing and Philosophy* (NA-CAP'08), Indiana University, Bloomington, Indiana, and the *Fifth European Conference on Computing and Philosophy* (E-CAP'07), University of Twente, Netherlands. I am grateful to the participants in these meetings for their helpful discussions. In particular, I would like to acknowledge the help, useful comments and criticisms of the members of the IEG, in particular Luciano Floridi and Matteo Turilli.

References

Alpern, K. D. (1997). What Do We Want Trust to Be? *Business and Professional Ethics Journal of Philosophy, Special Issue on Trust and Business: Barriers and Bridges, ed. by D.*

Koehn, 16(1-3), 29–46.

Archetype/Sapient, C. R. a. S. (January, 1999, http://www.gii.com/ trust_study.html). eCommerce Trust Study (Publication. Retrieved May 12, 2000, from monograph:

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Ba, S., Whinston, A. B., & Zhang, H. (1998). *Building Trust in the Electronic Marke through an Economic Incentive Mechanism.* Paper presented at the Proceedings of the International Conference on Information Systems, Charlotte, NC.

Baier, A. (1995). Trust and Its Vulnerabilities. In *Moral Prejudices: Essays on Ethics*. Cambridge, MA: Harvard University Press.

Berg, T. C., & Kalish, G. I. (1997). Trust and Ethics in Employee-Owned Companies. *Business and Professional Ethics Journal of Philosophy, Special Issue on Trust and Business: Barriers and Bridges, ed. by D. Koehn*, 211–224.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34–43.

Bhattacherjee, A. (2002). Individual Trust in Online Firms: Scale Development and Initial Test. *Journal of Management Information Systems, 19*(1), 211-241.

Blaze, M., Feigenbaum, J., & Lacy, J. (1996). *Decentralized Trust Management.* Paper presented at the IEEE Symposium on Security and Privacy, Oakland, CA.

Brenkert, G. (1998). Trust, Business and Business Ethics: An Introduction. *Business Ethics Quarterly, 8*(2), 195-203.

Cantrell, S. (2000). E-Market Trust Mechanisms. *Accenture Research Note: E-Commerce Networks, 11*, 1-3.

Castelfranchi, C., & Falcone, R. (1998). *Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification* Paper presented at the Proceedings of the Third International Conference on Multi-Agent Systems.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies 58*(6), 737-758

Dribben, M. R. Exploring the Processual Nature of Trust and Cooperation in Organisations: A Whiteheadian Analysis. *Philosophy of Management. Special Issue on Organisation and Decision Processes, ed. L. Leonard Minkes and T. Gear, 4*(1), 25–39.

Flores, F. L., & Solomon, R. C. (1997). Rethinking Trust. *Business and Professional Ethics Journal. Special Issue on Trust and Business: Barriers and Bridges,*

ed. D. Koehn, 16(1-3), 47–76.

Floridi, L., & Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines, 14*(3), 349-379.

Foster, I., & Kesselman, C. (Eds.). (1998). *The Grid, Blueprint for a New Computing Infrastructure*: T. Morgan

Kaufmann Inc.

Fukuyama, F. (1998). *The Virtual Handshake: E-Commerce and the Challenge of Trust*. Paper presented at the The Merrill Lynch Forum.

Gambetta, D. (1998). Can We Trust Trust? In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 213–238). Oxford: Basil Blackwell.

Gefen, D. (2000). E-Commerce: The Role of Familiarity and Trust. *Omega 28*(6), 725-737

Hosmeh, L. T. (1995). Trust: The Connecting Link between Organizational Theory and Philosophical Ethics. *Academy of Management Review, 20*(2), 379-403.

Jones, K. (1996). Trust as an Affective Attitude *Ethics and Information Technology, 107*(1), 4-25.

Jøsang, A., Keser, C., & Dimitrakos, T. (2005). Can We Manage Trust? In P. Herrmann, V. Issarny & S. Shiu (Eds.), *Proceedings Trust Management, Third International Conference (iTrust 2005)* (pp. 93-107). Paris, France.

Luhmann, N. (1979). *Trust and Power*. Chichester: John Wiley.

Maes, P. (1990). *Designing Autonomous Agents*. Cambridge, MA: MIT Press.

Nissenbaum, H. (2001). Securing Trust Online: Wisdom or Oxymoron. *Boston University Law Review, 81*(3), 635-664.

Oram, A. (2001). *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. Sebastopol, CA: O'Reilly & Associates Inc.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different after All: A Cross-Discipline View of Trust. *Academy of Management Review, 23*(3), 393-404.

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.

Rutter, J. (2001). From the Sociology of Trust towards a Sociology of "E-Trust". *International Journal of New Product Development & Innovation Management 2*(4), 371-385

Schmitt, F. (1987). Justification, Autonomy, Sociality. *Synthese, 73*, 43-85.

Solomon, R. C., & Flores, F. (2001). *Building Trust in Business, Politics, Relationships, and Life*. Oxford: Oxford University Press.

Stahl, B. C. (2004). *The Problem of Trust Creation through Technical Means in E-Commerce.* Paper presented at the Interdisciplinary Corporate Social Responsibility Conference, Nottingham University.

Stahl, B. C. (2006). *Trust as Fetish: A Critique.* Paper presented at the Journal of Information, Communication, Society 10th Anniversary International Symposium, York.

Tuomela, M., & Hofmann, S. (2003). Simulating Rational Social Normative Trust, Predictive Trust, and Predictive Reliance between Agents. *Ethics and Information Technology, 5*(3), 163-176.

Weckert, J. (2005). Trust in Cyberspace. In R. J. Cavalier (Ed.), *The Impact of the Internet on Our Moral Lives* (pp. 95-120). Albany: University of New York Press.

Wooldridge, M. (2002). *An introduction to multiagent systems*. Chichester: J. Wiley.

Wooldridge, M., & Jennings, N. R. (1995). Agent Theories, Architectures, and Languages: a Survey. In M. Wooldridge & N. R. Jennings (Eds.), *Intelligent Agents* (pp. 1-22). Berlin: Springer-Verlag.

[1] For more details the interested reader is referred to: for business ethics, (Alpern, 1997), (Solomon & Flores, 2001), (Berg & Kalish, 1997), (Brenkert, 1998), (Flores & Solomon, 1997); for e-commerce, see (Ba, Whinston, & Zhang, 1998), (Cantrell, 2000), (Fukuyama, 1998), (Gefen, 2000), (Stahl, 2004), (Stahl, 2006); and for system management, see (Bhattacherjee, 2002), (Dribben), (Hosmeh, 1995),

[2] See for example (Rousseau, Sitkin, Burt, & Camerer, 1998) and (Rutter, 2001)

[3] A similar approach has been also described in (Jones, 1996).

[4] (Baier, 1995)

[5] (Archetype/Sapient, January, 1999, http://www.gii.com/ trust_study.html http://www.gii.com/ trust_study.html #77 http://www.gii.com/ trust_study.html #77) and (Corritore, Kracher, & Wiedenbeck, 2003).

[6] For a more in depth description of the debate on the definition of AAs see (Russell & Norvig, 1995), (Russell & Norvig, 1995), (Maes, 1990), (Schmitt, 1987)

[7] In this and in the following sections I will use the words trust and e-trust in the same way they are used in the literature provided in this paper. Trust and e-trust will mean a decision taken by an AA that determines the circumstances of its actions. The parameters according to which this decision is taken, and its effects are specified time by time following the different analyses reviewed. In the same way the distinction between trust, trustworthiness, calculative trust and initial trust will be mentioned only when they appear in the reviewed analysis.

[8] See section **E-trust in artificial distributed systems: AAs** for more details about the debate on the definition of AAs.

[9] In this section I follow the terminology used by the authors, and refer to 'normative trust' and not to normative e-trust. This account concerns trust in MAS and seeks to explain occurrences of e-trust.