Trust as Individual Asset in a Network: A Cognitive Analysis

Francesca Marzo and Cristiano Castelfranchi

Abstract One of the most addressed kinds of relationships able to generate value in a network is trust (studied in different ways, for different purposes, by different disciplines). While it is very important to keep on investigating this issue using an interdisciplinary and integrated approach, it is also crucial to study the value-generation involved from a perspective able to disentangle what happen both at macro and micro level. In other terms we need a clear distinction between what is commonly called social capital and what is better known as relational capital. In this chapter we focus on the latter by analysing the point of view of the trustee on the bases of the goal oriented theory of trust. We developed a cognitive model of trust to explain why it represents a form of power and, then, a strong asset for individuals acting in a network. Finally, we propose some experimental future works based on this analysis and involving both laboratory experiments and multi-agents systems simulations.

Keywords Trust · Relational capital · Cognitive analysis · Multi-agent systems

1 Introduction

The next generation of pervasive and adaptive ICT applications will need to be *designed for trust*, to fulfill their expected impact on future Information Society [1]. Trust-oriented technologies imply a two-pronged challenge. First, pervasive and adaptive applications need to elicit adequate levels of trust in users, that must be willing to rely on largely autonomous devices and self-organized systems

F. Marzo (✉)
CeRSI, LUISS Guido Carli, Rome, Italy

C. Castelfranchi
ISTC-CNR, Rome, Italy

(users' trust) [2, 3]. Second, system components need reliable trust heuristics for autonomously coordinating with each other, in order to strike an optimal balance between flexibility and dependability (trust technologies). Hence this second basic condition asks for a detailed formal and computational model of trust that will be integral part of the digital pervasive environments of the future [4] but that will be also tractable for the computational units embedded in the environment. These distributed devices in fact must be designed for trusting each other despite the volatility of their interaction situations, the uncertainty intrinsic to real environments, the novelty of the contexts and of the possible peers but given their limited computational power. In this perspective, trust is a crucial decision criterion for agents with finite resources: it enables agents to decide whether to perform actions or delegate activities, evaluating risks and utilities. According to different perceived trust levels agents may adopt the proper policies, using different strategies. But to serve this purpose, the metrics and methods used for trust assessment must prove reliable also in highly dynamic and largely unknown environments [5]. On the contrary, current computational trust models are usually built either on the agent's direct experience of an interacting partner (interaction trust) [6, 7], or on reports provided by third parties about their experiences with a partner (witness reputation) [8–10]. Both these models present critical drawbacks when they need to cope with open, highly dynamic environments. Whenever starting a new session, for instance, agents are often engaged with totally unknown peers, and they must decide whether to trust them or not in a situation of partial ignorance, high uncertainty, and lack of prior knowledge: in such conditions neither witness reputation nor interaction trust provide significant guidelines for the agent's decisions [11]. It is important to overcome these critical shortcomings of existing models and applications of trust in computational systems, by investigating the critical cognitive capabilities needed for a quick and effective process of trust formation in open environments [5]. Our study is a step on this direction. One of the main objectives of our cognitive approach, in fact, is to build artificial agents which are able to reason about trust, that is, to improve actual computational and formal models of trust in online interaction between autonomous artificial agents and in human–machine interactions in such a way that they can account for sophisticated social processes of trust formation, revision, attribution, and circulation in open environments. In order for artificial agents to be able to take into account the complexity and multi-factoriality of trust they need to be built as agent able to find adaptive strategies and integrate different features for trust attribution and inducement. This will lead to a step change in computational, mathematical and logical models of trust for multi agents systems, ubiquitous computing, and human–machine interaction.

In this preliminary study we will focus on a specific side of trust relationship, the trustee, to reach two different aims: on one side, we will underline the intrinsic importance of trust links in a network, on the other, we will propose a theory that can explain how these links, that represent a real asset for nodes in the network, can be manipulated in order to accumulate power.

2 Social Capital and Relational Capital: Different Assets for Different Beneficiaries

Trust is sometimes a property of an environment, rather than of a single agent or even a group: under certain conditions, the tendency to trust each other become diffused in a given context, more like a sort of acquired habit or social convention than like a real decision [2, 12]. These processes of 'trust spreading' are very powerful in achieving high level of cooperation among large population, and should be studied in their own right. In particular, it is crucial to understand the subtle interaction between social pressures and individual factors in creating these 'trusting environments', and to analyze both advantages and dangers of such diffused forms of trust. In fact, considering collectivity and individuals as two different stakeholders, it is possible to say that in building trust their goals can, under certain circumstances, not only be different, but also contradictory: while the population of agents as a whole might gain more benefits from a spread diffusion of trust in the society as a whole, individual agents might seek for concentrating trust relationships on itself. In the first case we are dealing with what we commonly call "social capital", in the second one we are talking of what we call "relational capital".

The notion of social capital suggests an abstract hidden resource, which can be accumulated, tapped, attained when people value relationships among each other, interact, collaborate, learn and share ideas. This is a valuable stock of capital. Productive resources can reside not just in things but also in social relations among people [13, 14]. Resnick [15] argues that social capital is a residual side effect of social interaction and the enabler of future interactions. Brehm and Rahn [16] have developed a structural model that shows how social capital manifests itself in individuals as a relationship between levels of civic engagement and interpersonal trust. Starting from the assumption that if interpersonal trust increases then also civic engagement increases, there is a common agreement that we need to address the issue of trust to study the origin of social capital. It is possible to consider separately what are the connections between people (communities' fundamental characteristic) and what is the added value of these connections (economic aspect of social capital). These practices foster powerful norms of generalized reciprocity: the implication is that past collaboration is the basis for future collaboration, and refusal to take or give increases one's chances of being sanctioned or even removed from the society [17]. Hence social capital is essential for both personal and community development in the society. Nevertheless, individuals can also use their capital of trust for anti-social purposes. In order to study the individual form of this capital and then to start understanding how the two can be in contrast, we believe we need to analyze what it means that trust represents a strategic resource for agents that are trusted, proposing a model of 'trust as a capital' for individuals and suggesting the implication for strategic action that can be performed. Our thesis is that to be trusted: (1) increases the chance to be requested or accepted as a partner for exchange or cooperation; (2) improves the

'price', the contract that the agent can obtain. Since the term "capital" refers to a commodity itself used in the production of other goods and services and the adjective "social" is used to claim that a particular capital not only exists in social relationships but also consists in some kind of relationships between subjects, it is clear that for the capital goods metaphor to be useful, the transformative ability of social relationships to become a capital must be taken seriously. This means that we need to start by finding out what is the competitive advantage not simply of being part of a network, but more precisely of being trusted in that network.

3 Cognitive Model of Trust Network

Trust is a multi-factorial and highly dynamic notion: it depends on many concurrent factors, it changes over time due to a variety of reasons, and it involves both specific mental states, cognitive capacities, and characteristic social attitudes and relations [18]. As a corollary, the influence of a single factor on trust dynamics is rarely linear: a given relevant feature (e.g., high competence) often fails to affect trust attribution, or do so in indirect and complex ways, due to interference of other significant elements (e.g., lack of motivation). In order to account for the intrinsic complexity and dynamicity of this crucial notion, we need to adopt a multi-factorial theory of the formation, revision, attribution, and circulation of social trust [5]. Such a theory is important not only to face a major conceptual challenge, but also to develop ground-breaking technologies addressing important priorities in the present societal and economical context, that will become even more pressing in the future. The new generation of distributed, agent-oriented systems requires the evolution of capabilities for cognitive and social interaction in "open" environment and systems where agents can freely join and leave at any time, and where the agents are owned by various stakeholders with different aims and objectives [19]. The aim of achieving robust social interaction is thus especially challenging in open environments, like the web (with its serious problem of unknown possible partners), virtual social spaces, and physical environments inhabited by many distributed 'intelligences' and agents. To face this challenge, we need to cope with the dynamicity of trust, by defining its lifecycle within the agent mind as part of the entire social system. We need to individuate the rational foundations of trust attribution, along with internal dynamic of trust formation and its relations with: (1) experiences of prior actions and outcomes of previous interactions; (2) communication and perception of specific "markers of trust"; (3) shared and/or certified reputation; (4) reasoning, i.e., analogy, deduction; (5) transferability of trust attribution within and between different domains.

A "trust network" is the network of the trust relationships among several agents [5]. Each node is the source of possible trust attitudes and acts towards other agents, but it is also the trustee, which means that it receives several trust attitudes/ evaluations (and potential trust acts and relationship). This create a specific topology of the net that can assume a very centralized shape (converging net) or a

quite decentralized one. If a node is very central with many afferent links, the effect on the network will be greater, by affecting many nodes, than if it is a local and marginal node. More specifically, the trust links towards a node can have alternatives (being in an "or" relation with other links) and, then, generates competitors in the network [20].

We say that a node can trust other nodes for different domains and performances (independent trust links), or for the same task, as possible alternatives (alternative trust links), or a mixed form: one node can trust other nodes for given tasks (their co-power and coordination) and need all of them for the same outcome (interdependent trust links). Although it is important to consider such structural differences, what needs to be represented and managed are also the specific semantics of trust links, together with their intentionality and arguments. The trust relationship between nodes is then relative to a specific task (action, performance, service etc.) for a given goal in given context of the action. Trust dynamics and structural relations are affected by those multiple dimensions and arguments. For example, if y and z are possible referents (trustees) of x for the same need and task they are in competition with each other, and the crisis of y's trustworthiness might affect very much the relationship of x towards z. On the contrary, if x trusts y for a given thing and z for a different thing, the disappointment towards y will not necessarily affect x–z relationship [21]. Moreover, trust (which is based on beliefs of different kinds: evaluations, expectations, attribution beliefs, dependence beliefs) has certain "sources"; those beliefs are credible because of the sources they derive from. From what has been said so far, it is clear that the trust network, build on tasks, goals etc., is basically connected with another network: the dependence one.

The theory of dependence includes two type of dependences: (1) The objective dependence, which says who needs who for what in a given society. This dependence has already the power of establishing certain asymmetric relationships in a potential market. (2) The believed dependence, which says who is believed to be needed by who. This dependence is what determines relationships in a real market and settles on the negotiation power. The importance of dependence network for negotiation power has already been proved: the bigger is the number of people who depend on me for a given goal and the smaller is the number of those I depend on, the bigger will be my negotiation power. But this model is incomplete, since, although it is important to consider dependence relationship between agents in a society, there will be not exchange in the market if there is not trust to enforce some connection. That is to say that if a node is strongly needed by other nodes, but not trusted, her negotiation power does not improve [20].

4 Trust as Relational Capital

Thanks to a structural theory of what kind of beliefs are involved it is possible not only to answer some very important questions about agents' power in network but also to understand the dynamical aspects of relational capital. In addition, it is

possible to study what a difference between trustee's beliefs and others' expectations on her implies in terms of both reactive and strategical actions performed by the trustee. First, let us consider what kind of strategies can be performed to enforce the other's dependence beliefs and his beliefs about agent's competence. Since dependence beliefs is strictly related with the possibility of the others to see the agent in the network and to know her ability in performing useful tasks, the goal of the agent who wants to improve her own relational capital will be to signaling her presence and her skills. While to show her presence she might have to shift her position (either physically or figuratively like, for instance, changing her field), to communicate her skills she might have to hold and show something that can be used as a signal (such as certificate, social status etc.). This implies, in her plan of action, several and necessary sub-goals to make a signal. This sub-goals are costly to be reached and the cost the agent has to pay to reach them can be taken has the evidence for the signals to be credible (of course without considering cheating in building signals). It is important to underline that using these signals often implies the participation of a third subject in the process of building trust as a capital: a third part which must be trusted. We would say the more the third part is trusted in the society, the more expensive will be for the agent to acquire signals to show, and the more this signals will work in increasing the agent's relational capital. We will see later how this is related with the process of transferring trust from an agent to another (building reputation).

Let us now consider how willingness beliefs can be manipulated. In order to do so, consider the particular strategy performed to gain the other's good attitude through gifts. It is true that the expected reaction will be of reciprocation, but this is not enough. While giving a gift the agent knows that the other will be more inclined to reciprocate, but she also knows that her action can be interpreted as a sign of the good willingness she has: since she has given something without being asked, the other is driven to believe that the agent will not cheat on him. Then, the real strategy can be played on trust, sometimes totally and sometimes only partially—this will basically depend on specific roles of agents involved. On the other hand, relational capital can also decrease. Losing relational capital means to be discredited and it can be imputed to the fact that some of the strategies performed to make the others trust fail. In fact it is possible that if the goals of signaling competences fail to be reached (because the signs chosen is bad, for instance) it is not necessarily true that the agent will just not increase her relational capital, but she can also lose some (since the agent who should trust her can valuate the sign particularly badly, for some reason). Also, if the tentative to enter the dependence network of some agents does not get true, it could be the case of losing relational capital in another market, both for the effort put in the action, which is time consuming, and for the fact that agents in the existing network can feel "betrayed". Finally, if the agent's attempt to show her willingness is interpreted as opportunistic exchange, the agent who was supposed to trust her can react badly and harmed her reputation. Another important feature of the dynamic of relational capital is the possibility of transferring from agent to agent. In fact, relational capital can also circulated inside a given society. If somebody has a good

reputation and is trusted by somebody else, she can be sure this reputation will pass and transfer to other actors—and this is always considered in marketing strategies. What is not clear yet is how these phenomena work. But when trust on an agent propagates, it is strategically important for the agent to know very well how this happens and which ways trust takes to expand. That said, it should be clear the importance of understanding if and how much an agent is able to manage this potentiality of her capital, also taking into account the fact that there might exist several type of discrepancies in subjective valuation (i.e., differences between how the others trust an agent and the level of trustworthiness that agent perceive in herself) and that these discrepancies can deeply influence in terms of strategical actions that can be performed.

5 Conclusion

Online interactions can be described as embodied in three foundational pillars (other than the technological one): regulative, normative, and cognitive [22]. Regulative aspects are thought to be based upon legal sanction. Normative aspects are morally grounded and people will comply with these elements based on social obligation. Cognitive aspects are individual mental states that can be assumed or modeled and that contribute to the collective constructions of social reality via meaning systems and other rules. It is this cognitive aspect that emphasizes the taken for granted beliefs to which individuals will conform. The different interaction conceptions should be viewed as frames for understanding the underlying tensions in online society rather than independent categories or alternatives. Our focus in this chapter has been on cognitive aspects of interactions and on a particular concept that has been recognized as crucial for any kind of interaction between individuals: trust. This concept has been extensively studied in several disciplines and, in particular, in the context of e-services the focus has been mostly on how trust affects users' intention to buy or re-use online services and on designing computational trust models to predict degrees of trust. The aim of our study is, instead, to study these processes in an environment constituted by heterogeneous agent (being it a condition already diffused but that will be the very standard in future) the models developed following the cognitive analysis presented in this chapter are intended to be tested on empirical bases (e.g. through experiments aimed to verify the actual strategies applied by humans to manipulate their relational capital under different circumstances, or collecting evidences in the interaction between human agents and artificial agents). At the same time, future work will produce advanced agent-based test-beds and social simulations, to test and verify not only interaction between artificial agents but also whether systems endowed with different simple strategies/heuristics for trust assessment and inducement in a trust network are better in performance than systems that manipulate trust only on the basis of previous experiences and/or reputational information. Both this test can be run by using a very interesting framework

already used in multi-agent systems: Colored Trails [23–25]. It is in fact an open source very adaptable framework designed to be used both for laboratory experiments (with the possibility of using tasks more complex than those usually used in laboratory investigation about interaction) and for agent-based simulation [26].

References

1. Ramchurn, S.D., Huynh, D., Jennings, N.R. (2004). Trust in multi-agent systems. *The Knowledge Engineering Review 19*(1), 1–25.
2. Ricci, A., & Omicini, A. (2004). Engineering trust in complex system through mediating infrastructures. WOA.
3. Castelfranchi, C., & Tan, Y. H. (Eds.). (2001). *Trust and deception in virtual societies*. Dordrecht: Kluwer Academic Publisher.
4. Kagal, L., Finin, T., & Joshi, A. (2001). Trust based security for pervasive computing environments. *IEEE Communications, 34*(2), 154–157.
5. Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and Computational model*. UK: Wiley.
6. Witkowski, M., Artikis, A., & Pitt, J. (2001). Experiments in building experiental trust in a society of objective-trust based agents. In R. Falcone, M. Sing, & Y.-H. Tan (Eds.), *Trust in cyber-societies*. Berlin: Springer.
7. Jonker, C. M., & Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In F. J. Garijo & M. Boman (Eds.), *Multi-agent system engineering: Proceedings of the 9th European workshop on modeling autonomous agents in a multi-agent world, lecture notes in artificial intelligence* (Vol. 1647). Berlin: Springer.
8. Carbó, J., Molina, J. M., & Dávila, J. (2005). Fuzzy referral based cooperation in social networks of agents. *AI Communications, 18*, 1.
9. Yu, B., Singh, M.P., Sycara, K. (2004). Developing trust in large-scale peer-to-peer systems. In *Proceedings of the First IEEE Symposium on Multi-Agent Security and Survive-ability*.
10. Sabater, J., & Sierra, C. (2005). Review on computational trust and reputation models. *Artificial Intelligence Review*, 24 (1).
11. Billhardt, H., Hermoso, R., Ossowski, S., Centeno, R. (2007). Trust-based service provider selection in open environments. In *proceeding of the he 22nd Annual ACM "Symposium on Applied Computing"*.
12. Arcos, J. L., Esteva, M., Noriega, P., Rodríguez, J. A., & Sierra, C. (2005). Engineering open environments with electronic institutions. *Journal on Engineering Applications of Artificial Intelligence, 18*(2), 191–204.
13. Coleman, J.S (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94.
14. Putnam, R. (1993). The prosperous community: social capital and public life. In American Prospect, Vol. 13.
15. Resnick, P. (2002). Beyond bowling together: Sociotechnical capital. In J. M. Carroll (Ed.), *HCI in the new millennium*. New York: Addison-Wesley.
16. Brehm J., & Rahn W. (1997). Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science*, 41(3).
17. Bowles S., & Gintis H. (2002). Social capital and community governance. *Economic Journal*, 112.
18. Falcone, R., Singh, M., & Tan, Y. H. (2001). *Trust in cyber-societies: Integrating the human and artificial perspectives, LNAI 2246*. Berlin: Springer.
19. Wooldridge, M. (2002). *An introduction to multi-agent systems*. Chirchester: Wiley.

20. Castelfranchi, C., Falcone, R., & Marzo, F. (2007). *Being trusted in a social network: Trust as relational capital, Lecture notes in computer science*. Germany: Springer.
21. Castelfranchi C. (2012). Earthquakes in trust networks: Basic dynamic principles (3 Apr 2012) SSRN Working Papers Series.
22. Scott, W. R. (1995). *Institutions and organizations*. CA: Sage.
23. Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., & Shieber, S. (2010). Agent decision-making in open mixed networks. *Artificial Intelligence, 174*(18), 1460–1480.
24. Gal Y., Pfeffer A., Marzo F., Grosz B. (2004). Learning social preferences in games. In *proceeding of the National conference on artificial intelligence* (AAAI). San Jose: California.
25. Grosz B.J., Kraus S., Talman S., Stossel B., Havlin M. (2004). The influence of social dependencies on decision-making: Initial investigations with a new game. In *Third International Joint Conference on Autonomous Agents and Multi-Agent Systems* (pp. 782–789). New York.
26. Gal Y., Grosz B., Kraus S., Pfeffer A., Shieber S. (2005). Colored Trails: a Formalism for Investigating Decision-making in Strategic Environments. In *proceeding of the IJCAI Workshop on Reasoning, Representation, and Learning in Computer Games* (pp. 25–30). Edinburgh, Scotland, August 2005.