Ethics Inf Technol (2011) 13:1–3
DOI 10.1007/s10676-010-9263-1

The case for e-trust

Mariarosaria Taddeo · Luciano Floridi

Published online: 9 January 2011

Trust is generally understood as a relationship in which an agent (the trustor) decides to depend on another agent's (the trustee) foreseeable behaviour in order to fulfil his expectations. It is a fundamental aspect of social interactions, as it has economical, social, psychological and ethical implications, and as such it is a crucial topic in several research areas (Gambetta 1998; Taddeo 2009).

In the past decade, research interest in trust-related concepts and phenomena has escalated following the advent of the information revolution (Floridi 2008). The use of Information and Communication Technologies, of Computer Mediated Communications (CMCs), and the development of artificial agents—such as SatNav systems, drones, and robotic companion—have provided unprecedented opportunities for social interactions in informational environments, involving human as well as artificial and hybrid agents (Ess 2010). In such scenario, one of the most problematic issues is represented by the emergence of e-trust, that is, trust specifically developed in digital contexts and/or involving artificial agents. Like trust, e-trust too has a variety of heterogeneous implications, ranging from the effects on social interactions in digital environment to the behaviour of the involved agents, whether human or artificial (Taddeo 2009; Ess 2010).

M. Taddeo (✉) · L. Floridi
University of Hertfodshire, Hatfield, UK
e-mail: m.taddeo@herts.ac.uk

L. Floridi
e-mail: l.floridi@herts.ac.uk

M. Taddeo · L. Floridi
University of Oxford, Oxford, UK

When e-trust is considered from a philosophical perspective, four problems seem to be more salient: (i) the identification of the fundamental and distinctive aspects of e-trust; (ii) the relation between trust and e-trust, that is, whether e-trust should be considered as an occurrence of trust on-line or as an independent phenomenon in itself; (iii) whether the environment of occurrence has any influence on the emergence of e-trust; and, finally, (iv) the extent to which artificial agents can be involved in an e-trust relationship.

Problems (i)–(iv) are usually addressed together (Johnson 1997; Nissenbaum 2001; Weckert 2005). The literature focuses on whether the characteristics of on-line environment and social interactions satisfy the minimal requirements for the emergence of trust. It is often argued that two conditions are necessary to this purpose: (a) the presence of a shared cultural and institutional background, and (b) certainty about the trustee's identity. The debate on the possibility of e-trust leads then to two opposite views: some argue that conditions (a)–(b) cannot be met in the on-line environment (Pettit 1995; Seigman 2000; Nissenbaum 2001), and therefore that there cannot be e-trust; others argue that they can (Weckert 2005; Vries 2006; Papadopoulou 2007; Turilli et al. 2010) and hence that e-trust is possible.

The analysis of (iv) depends upon the role attributed to the feelings and the psychological status of the agents involved in a trust relation. Those who deem feelings and emotion necessary for the occurrence of trust deny the possibility that trust (including the online variety) may ever be present when one of the two peers is an artificial agent (Jones 1996). The opposite thesis is defended by those who consider (e-)trust a phenomenon occurring independently from the emotional and psychological status of the involved agents (Taddeo 2010).

This special issue presents five articles, which propose original analyses addressing problems (i)–(iv). All articles share a *hybrid approach*, as they all investigate e-trust, balancing the conceptual analysis of the features of this phenomenon with the consideration of either empirical data from lab experiments or models developed in AI.

The issue has the twofold goal of providing the reader with cutting edge research on e-trust while also promoting the use of the hybrid approach for its analysis, and more generally for the analysis of a plethora of other phenomena concerning the interactions of human and artificial agents in the informational environment—such as, for example, informational warfare, e-life, cyber- terrorism and crime. All these phenomena depend on the context of their occurrence, the implementation of the given technologies, the features of the involved agents and the dynamics of their interactions. With this respect, empirical data and AI models provide both the grounding for studying such phenomena and the means to test the accuracy of the developed conceptual analysis.

Bicchieri's and Lev-On's article opens the special issue by studying the effects of CMCs on the emergence of cooperation and trust. The article argues that the processes of trust-building and maintenance are influenced more by pre-play communication and group size than by communication medium, and provides the results of laboratory experiments in support of such argument.

Grodzinsky's, Miller's and Wolf's paper offers a comprehensive review of the literature on e-trust. It also proposes an original perspective on the analysis of this phenomenon, by taking into consideration the role of the *humans* who design, implement and deploy the artificial agents involved in the e-trust relation. The article develops an object-oriented model for the analysis of the different kinds of occurrences of trust.

Simpson's contribution focuses on the concept of reputation and its assessment in on-line interactions. Reputation is considered a core parameter in the evaluation of the trustee's trustworthiness, and therefore the accuracy of such an evaluation is believed to be of paramount importance for the occurrence of e-trust. The paper first analyses some of the problems that on-line reputation systems face and then proposes eleven principles for the design of such systems.

Tavani's and Buechner's analysis aims to show that conceptual analyses of trust benefit from the modelling of trust relations developed in AI using distributed systems. The article endorses Walker's definitions of *diffuse trust* and *default diffuse trust* (Walker 2006) to study cases of commitment and trust among the artificial agents of a distributed system. Such cases are studied in the light of the experiment conducted with SPIRE (Shared Plans Intention Reconciliation Experiments).[1] The authors show how the data from the experiments help to understand trust and the dynamics of its occurrences, as well as unveiling important ethical issues.

Finally, Pieters' article considers the issue of e-trust toward artificial systems, and relates the trust that users may have in such systems to the *explanation* provided to the users concerning the systems' function. He argues that the relation between explanation and trust is critical in cases of e-trust, as the explanation provides the means for assessing the system's trustworthiness. A distinction is drawn between *explanation-for-trust* and *explanation-for-confidence*. The former is the explanation of how a system works, by describing details of its internal operations, while the latter is an explanation aiming at making the user feel comfortable in using the system.

Before leaving the reader to the articles of this special issue, we would like to express our gratitude to all the contributors for their collaboration during every phase of this project. We believe their work provides an enlightening contribution to the understanding of the subject of e-trust and to the development of the debates surrounding it. Finally, we thank Jeroen van den Hoven and Springer for their outstanding support and assistance while we were working on this project.

References

Ess, C. M. (2010). Trust and new communication technologies: vicious circles, virtuous circles, possible futures. *Knowledge Technology and Policy, 23*(3/4), 287–305.

Floridi, L. (2008). Artificial intelligence's new frontier: Artificial companions and the fourth revolution. *Metaphilosophy, 39*(4/5), 651–655.

Gambetta, D. (1998). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–238). Oxford: Basil Blackwell.

Johnson, D. G. (1997). Ethics Online, shaping social behavior online takes more than new laws and modified edicts. *Communication of the ACM, 40*(1), 60–65.

Jones, K. (1996). Trust as an affective attitude. *Ethics and Information Technology, 107*(1), 4–25.

Nissenbaum, H. (2001). Securing trust online: Wisdom or oxymoron. *Boston University Law Review, 81*(3), 635–664.

Papadopoulou, P. (2007). "Applying virtual reality for trust-building e-commerce environments". *Virtual Reality, 11*(2–3), 107–127.

Pettit, P. (1995). "The cunning of trust." *Philosophy & Public Affairs, 24*(3), 202–225.

Seigman, A. B. (2000). *The problem of trust*. Princeton, NJ: Princeton University Press.

Taddeo, M. (2009). Defining trust and e-trust: Old theories and new problems. *International Journal of Technology and Human Interaction (IJTHI), 5*(2), 23–35.

[1] SPIRE is an experimental system that models the intention of reconciliation and commitment of artificial agents in the context of collaborative activities.

Taddeo, M. (2010). Modelling trust in artificial agents, a first step toward the analysis of e-trust. *Minds and Machines, 20*(2), 243–257.

Turilli, M., Vaccaro, A., et al. (2010). "The case of on-line trust". *Knowledge Technology and Policy, 23*(3/4), 333–345.

Vries, P. d. (2006). Social presence as a conduit to the social dimensions of online trust. In W. IJsselsteijn, Y. d. Kort, C. Midden, B. Eggen, & E. v. d. Hoven (Eds.), *Persuasive technology* (pp. 55–59). Berlin/Heidelberg, Springer

Walker, M. U. (2006). *Moral repair: Reconstructing moral relations after wrongdoing*. New York: Cambridge, Cambridge University Press.

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.