

Trust in Technology: a Distinctive and a Problematic Relation

Mariarosaria Taddeo
Department of Philosophy
University of Hertfordshire
Uehiro Centre
University of Oxford
United Kingdom
m.taddeo@herts.ac.uk

The use of tools and artefacts is a distinctive and problematic phenomenon in the history of humanity, and as such it has been a topic of discussion since the beginning of Western culture, from the myths of the Ancient Greek, through Humanism and Romanticism to Heidegger. Several questionable aspects have been brought to the fore: the relation between technology and arts, the effects of the use of technology both on the world and on the user, and the nature of the trust that users place in technology (Mitcham 2003). This last topic is the subject of this special issue, which has the twofold goal of fostering a cross-disciplinary debate and, in doing so, of overcoming, at least in part, the fragmentation of the literature on this topic.

The problematic nature of trust in technology becomes evident with the dissemination of information and communication technologies (ICTs) and the subsequent information revolution, with which artefacts cease to be used mainly to perform physical and fatiguing tasks, and begin to be deployed to execute also intellectual works (Floridi 2008). As the outsourcing to (informational) artefacts becomes more pervasive, the *trust* and the dependence of the users on such artefacts also grows, bringing to the fore fundamental questions concerning some of the most relevant debates in philosophy of technology, such as the definition of technology, the criteria that should be endorsed to determine whether an artefact is trustworthy, and those that should guide the design of artefacts. At the same time, the analysis of trust in technology draws attention to issues debated in philosophy, ethics and computer ethics, like the nature of trust, the necessary requirements for its occurrence, whether trust can be developed toward an artefact or can only concern human beings and, finally, whether trust can be present in digital environments (Taddeo 2009).

The articles presented in this special issue address these questions while embracing different backgrounds. They are organised in three groups, corresponding to the three main aspects of the debate that they analyse: the occurrence of trust in

digital environments, the nature of trust in technology, and the relation between trust, technology and design.

In more detail, the articles of the first group form the first volume of the special issue concern problems posed by the emergence of on-line trust, the development of trust toward artificial agents, and trust and responsibility in on-line environments. Ess's article opens the volume, providing an enlightening analysis of the problems related to the occurrence of on-line trust and of the arguments pro and contra such occurrences. This article lays the ground for the other papers in this group.

Cheshire, Antin, Cook and Churchill's paper provides a sociological analysis of the occurrences of on-line trust. The authors present the result of an on-line survey of Internet users and analyse the relation between two types of on-line trust - *general trust*, trust in most websites; and *familiar trust*, trust in websites that one frequently visits - as well as independent variables such as information technology competence, and adverse on-line events. On the basis of this survey, they conclude that users with some competence in ICT are less cautious in establishing general on-line trust, and that frauds or defections in on-line environments affects general trust but not familiar trust.

The analysis of the nature of on-line trust and of its ethical implications is developed in Turilli, Vaccaro and Taddeo's paper, where we analyse on-line trust on the basis of the definition of trust provided in (Taddeo 2010), and argue that on-line trust is an occurrence of trust that specifically qualifies the relation of communication among individuals in digital environments, and that on-line trust promotes the emergence of social behaviours rewarding honest and transparent communications.

Trust in digital environments also concerns the emergence of trust towards artificial agents as well as the concept of responsibility in such environments. The contribution by Durante offers an insightful analysis of the occurrences of trust towards artificial agents, so-called e-trust; while the issue of the responsibility for actions performed by trusted artificial agents is addressed by Pagallo's article. Cohen-Almagor's contribution closes the first group by addressing the important issues of trust and social responsibility on the Internet. The author analyses the responsibilities of Internet Service Provider (ISPs) and web-hosting companies and argues that both ISPs and web-hosting companies should take responsibility for content published and communicated through their channels.

The second group of papers opens the second volume of the special issue and focuses on the nature of trust in technology. Kiran and Verbeek's paper argues that technologies and technological artefacts help to constitute human subjectivity, and that trust in technology has the nature of confidence, as a user trusts herself *to* technology.

Nickel, Franssen and Kroes's article deepens the investigation of trust in technological artefacts. The authors distinguish between those occurrences of trust that are grounded on pure rational choice and what they call "motivation-attributing" occurrences of trust. They then argue that even in the cases of those artefacts which seem best suited to be trusted, such as intelligent systems and complex socio-technical systems, a motivation-attributing concept of trustworthiness cannot be straightforwardly applied. They therefore conclude that any applicable notion of trustworthy technology is significantly different from the notion of interpersonal trust.

Pitt's paper contributes to the debate on the nature of trust in technology by arguing that the question "Can we trust technology?" is not the question that should be asked, as it is ultimately open-ended. The author suggests that the only questions that should be asked, with respect to the trust placed in technology, are those that concern *specific* issues, which can then produce specific answers. Pitt concludes that trust in technology is actually a form of trust in the people who design and build technological artefacts.

The third and last group of articles investigates the relation between trust, technology and design. Selinger and Whyte's paper introduces the subject. In their article, the analysis of nudges provided in (Thaler and Sunstein 2008) is taken under consideration and it is argued that Thaler and Sunstein's advance the theory of how designers can improve decision-making in various situations where individuals have to make choices. The authors defend the thesis that the moral acceptability of nudges rests, at least in part, on whether they can provide an account of the competence required to offer nudges, an account which would serve to warrant our general trust in choice architects.

The other two papers presented in this group focus on the design of ICT artefacts. Vermaas, Tan, Hoven, Burgemeestre and Hulstijn's article considers the meaning, roles, and uses of trust in the economic and public domain, and focuses on the task of designing systems for trust in information technology. The authors analyse this task by means of a survey concerning what trust means in the economic and

public domain, using the model proposed by (Lewicki and Bunker 1995), as well as the paradigm of Value-Sensitive Design.

Finally, Monti's paper provides an interesting analysis of the importance of ethical choice in the design of informational artefacts. Monti argues that the intersection of trust, law and technology, as in the case of the Internet, can become either an empowering factor for businesses and individuals or a tool for infringing human rights. He stresses the importance of considering every technology as a human by-product, and that when technology fails, it is largely due to human fault.

Before leaving the reader to the articles of this special issue, I must express my gratitude to all the contributors for their collaboration during all phases of this project. I should also like to thank Luc Bovens, S. D. Noam Cook, Dorothy E. Denning, Michael R. Nelson, Trevor Pinch, and John Weckert, who kindly accepted the invitation to read and comment some of the papers presented in this special issue. I believe their work provides an enlightening contribution to our understanding of the subject of trust in technology and to the development of the debate that this special issue intends to promote.

I shall also express my gratitude to those colleagues who agreed to review the papers of this special issue; their efforts were fundamental to improving its quality. Finally, I shall thank Luciano Floridi and Springer for providing all the necessary support and assistance to develop and finish this project.

References

- Floridi, L. (2007). "A look into the Future Impact of ICT on Our Lives." The Information Society **23**(1): 59-64.
- Floridi, L. (2008). "Artificial intelligence's new frontier: Artificial companions and the fourth revolution." Metaphilosophy **39**(4-5): 651–655.
- Lewicki, R. J. and B. B. Bunker (1995). Trust in relationships: A model of trust development and decline. Conflict, cooperation and justice. B. B. Bunker and J. Z. Rubin. San Francisco, Jossey-Bass: 133-174.
- Mitcham, C. (2003). Three Ways of Being-With Technology. The Philosophy of Technology: The Technological Condition - An Anthology. R. C. Scharff and V. Dusek. Oxford, Blackwell: 490-507.
- 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.
- Taddeo, M. (2010). "Modelling Trust in Artificial Agents, a First Step toward the Analysis of e-Trust." Minds and Machines **20**(2): 243-257.