



HAL
open science

The Social Construction of “Shared Reality” in Socio-Technical Systems

Kristina Milanović, Jeremy Pitt

► **To cite this version:**

Kristina Milanović, Jeremy Pitt. The Social Construction of “Shared Reality” in Socio-Technical Systems. 12th IFIP International Conference on Trust Management (TM), Jul 2018, Toronto, ON, Canada. pp.149-159, 10.1007/978-3-319-95276-5_11 . hal-01855986

HAL Id: hal-01855986

<https://inria.hal.science/hal-01855986>

Submitted on 9 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The Social Construction of “Shared Reality” in Socio-Technical Systems

Kristina Milanović^[0000–0002–8918–9748] and Jeremy Pitt^[0000–0003–4312–8904]

Imperial College London
Exhibition Road, London, SW7 2BT, United Kingdom
{kristina.milanovic08, j.pitt}@imperial.ac.uk

Abstract. As the size, complexity and ubiquity of socio-technical systems increases, there is a concomitant expectation that humans will have to establish and maintain long-lasting ‘relationships’ with many types of digital artefact: for example with humanoid robots, driverless cars or software agents running on ‘smart’ devices. Rather than being limited to one-off interactions, these relationships will continue over longer time frames, correspondingly increasing the likelihood of errors occurring from numerous causes. When digital errors occur, often complete human mistrust and distrust is the outcome. The situation is exacerbated when the computer can make no act of reparation and no avenue of forgiveness is open to the human. In the pursuit of designing long-lasting socio-technical systems that are fit-for purpose, this position paper reviews past work in relevant social concepts and, based on the sociological theory of social constructivism, proposes a new approach to the joint human-computer construction of a “shared reality”.

Keywords: Trust, Forgiveness, Norms, Socio-Technical Systems, Social Constructivism

1 Introduction

In a world of increasing interconnectivity, socio-technical systems, which are defined as any multi-agent system where there is a mix of human and computational agents working together to achieve a common goal[1], are becoming increasingly prevalent in every day life. Agents can be human, physical computers (including robots) or digital (software). The digital or computational agents can be working with the human agents in any combination. This includes facilitating human-human communication/interaction via smart devices, a number of computational agents and human overseers such as a network of driverless cars or smart meters being monitored remotely and a human agent interacting with multiple digital agents in the form of digital home assistants in a smart home.

The ‘relationships’ here between the various computer, digital and human agents will not be one-off interactions, such as the purchase of an item on the internet with the help of a digital shopping assistant, where a single task needs to be accomplished and the agents either do not interact again, or do not interact until a long time period has passed. They will need to be maintained

through multiple interactions in order for the goal of the system to be achieved. Continuing with the preceding examples, this includes optimal energy pricing for neighbouring households in a smart energy network, safe operation of smart cars in traffic flow and efficient operation of devices within the home.

Like any device these computational and digital agents inherently have some level of error and humans occasionally make mistakes too. As the longevity and size of the system of agents increases, the likelihood of errors occurring also increases. This can be in the form of system faults or misunderstandings between the human and computational or digital agents. Since the agents in the system will need to interact over long time periods, agents will need to have mechanisms to enable these relationships to last over time in the face of these errors. One method of doing this is to apply concepts of trust and forgiveness which have enabled human society to maintain long-term relationships, to systems which include computational and digital agents. In order to use socio-cognitive theories, where trust is a mental state of a social cognitive agent (i.e.: an agent with goals and beliefs)[2], the non-human agents need to interact with the human agents in a manner akin to other humans.

In this position paper past work in computational models of trust, forgiveness, norms and values is reviewed. A new theory for how these concepts can be combined together addressing and how, specifically, the effect of system errors may be mitigated is proposed. The paper is split into two parts, the first section covers previous work on human relationship building mechanisms in computing while the second section outlines how the proposed new theory fits within this already established framework.

Pulling from sociology to apply utilise the trust and forgiveness mechanisms already well established in human society, the new proposal follows on from the theory of social constructivism. An artificial form of social constructivism is suggested which can be applied to socio-technical systems to create a “shared reality” based on shared experiences[3], in this case shared social norms and values. This will allow for the creation of more fit-for purpose models of socio-technical systems which should enable long-lasting ‘relationships’ between agents.

2 Social Concepts in Socio-Technical Systems

Socio-technical systems in particular are useful for their ability to model inherent human attributes which can have profound effects on the system. Concepts of trust, forgiveness and social norms have long been studied in sociology, psychology and evolutionary biology and their adaptation into the field of computing, to analyse network behaviour, has an advantage over simpler Game theoretic approaches as they are better able to take into account human attributes such as deception. Game theoretic models tend to simplify observed behaviour in order to model it, for example, they reduce the trust decision to a probability[2]. This form of modelling is insufficient when attempting to create a model that accurately represents human decision making, thus socio-cognitive approaches are better suited to it. This section covers previous work key human traits which

may be applied to socio-technical systems to assist with modelling and analysis and how they fit into the overarching framework of social capital.

2.1 Social Capital

Social capital encompasses concepts which are established as part of institutions of individuals such as trust, forgiveness and social norms and values. It was defined by Ostrom as “*an attribute of individuals that enhances their ability to solve collective action problems*” [4], where a collective action problem is one that requires the cooperation of multiple individuals to achieve a shared goal. Trust is the method, or “glue”, which enables the solving of these collective action problems[5].

Social capital can be broken down into three general categories: trustworthiness, social networks and institutions. The definition of institutions here is synonymous with the most generic of social norm definitions. According to Ostrom, institutions outline allowed and forbidden actions, or “*collections of conventional rules by which people mutually agree to regulate their behaviour*” [6]. Recent developments in the field have included the theory of electronic social capital[7], and upholds the idea that it can be used to solve collective action problems in a digital environment.[6]

Trust The concept of trust has been defined many times. From sociology, one often cited definition is that trust “*is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends*” [8]. Another definition, also from sociology, is that trust can be used to reduce complexity of social interactions [9]. Applying this definition to computational networks can be used to simplify behaviour in complex networks and it is this aspect in particular that is relevant to the proposal outlined here.

Computational trust has been a topic of increasing research over the past two decades since one of the first models was presented by Marsh in 1994.[10] Marsh proposed that incorporating trust into artificial intelligence would assist with social decision making by the agents. Following on from this, one of the earliest cognitive trust models[11] was outlined by Castelfranchi and Falcone[2] in 2001. They proposed that only agents with goals and beliefs, in other words cognitive agents, are capable of trust. Another proposal from Jones in 2002 suggested that trust has two components, a belief in another agent and the expectation that the agent would perform some action, or attain a desired goal[12].

Since then there have been a proliferation of trust models in multiple fields[13] where the concept of trust has which have been expanded to include extended definitions and situations. Concepts of mistrust (misplaced trust), distrust (how much the truster believes the trustee will consciously work against them in a given situation) and untrust (the truster does not believe the trustee will work with them in a given situation) have been postulated[14]. Specific models for use in areas as diverse as ecommerce systems[15] and multi-agent systems have developed. These models have been compared and contrasted at length in recent

years[11, 16]. Overall the key concept these models have in common is that there is a need for systems which mimic human social interactions, so that the agents are able to make decisions about who and how much to trust, without external input from human operators.

In terms of modelling, in large socio-technical systems, cognitive models of trust have advantages over more simplistic game theoretic models[11]. Game theory tends to assume that all system agents are rational, however this may not be the case in a system involving humans, where the agents may deceive each other or act in a (seemingly) irrational manner.[1] Models such as those proposed by Axelrod in the 1980's[17], are less able to represent the complicated relationships between agents when humans are involved.

This is particularly relevant when errors in the system occur. Errors can come in broadly two categories, malfunctions and mistakes. Malfunctions are where there is a bug in the code or a physical breakdown of some sort leading to agents not behaving in the expected manner. Mistakes occur where there has been some kind of misunderstanding between what the agent was expected to do and what they actually did. One way this has been addressed previously is through reputation mechanisms which aggregate past behaviour over time[18] or trusted third parties which act as a guarantee of trustworthiness[19]. The most visible application of this is in ecommerce, for example the in the form of ratings and reviews for online buyers and sellers.

Forgiveness When there has been a trust breakdown in human relationships, forgiveness is often used as a mechanism for rebuilding trust after a breakdown has occurred. It is not unconditional and is governed by a number of motivating factors. These factors include offence severity and frequency, as well as the offender's previous behaviour towards the victim, and the offender's intent (was there deliberate or accidental error) and subsequent efforts to reconcile or repair the relationship.[20] Forgiveness is a mechanism which allows the victim to replace their formerly negative feelings towards the offender with positive ones which would allow them to reconcile and maintain their relationship.[21] In fact, it may have evolved specifically in humans to allow valued social relationships to be maintained over time[22].

A socio-cognitive model, where social agents have cognisant goals and objectives, therefore may provide the basis for agents which are able to repair relationships, through forgiveness, when there has been a breakdown. This is something which can be seen even using game theoretic approaches. Axelrod found that in a series of iterated prisoner's dilemmas forgiving strategies were the most robust over time.[23] Additionally, in games where the structure of the agents was fixed, where agents had the same neighbours throughout the iterations, cooperative strategies dominated and agents were more forgiving and friendlier to their neighbours.[24]

Mechanisms for facilitating forgiveness have been proposed which allow offenders not only to apologise but also to validate their apology with some kind of reparation action.[25] Without reparations, it has been found that victims in

online interactions took revenge against the offender where their apology was not seen as costly enough by the victim.[25] Apologies need to be costly to the offender in order to facilitate forgiveness.[26] This serves two purposes as it punishes the offender for their offence and makes reparations towards the victim.

Another part of the forgiveness process is that the offender admits that they are guilty of a transgression. The victim is more likely to forgive the offender if they believe the offender is less likely to repeat the transgression in future interactions.[22] The acknowledgement of guilt itself is an intangible cost for breaking a social norm[27] and in face to face interactions, humans have emotional and physical reactions which shock the victim if they are regretting, or feeling ashamed or embarrassed by their actions, for example by blushing.[28] Since physical cues are difficult to transmit in socio-technical systems mechanisms enabling apology, admission of guilt and suitable reparations to the victim are key in order to facilitate forgiveness.

Social Norms and Values Although social norms and social values tend to be used interchangeably, social norms are specific guidelines which members of a society are expected to uphold in given situations, whereas social values are much broader[29]. They are the general ideas which a society holds true and include concepts such as trusting and forgiving each other[30].

Human interactions are governed by many social norms of behaviour which are learnt in childhood[31] and dictate what is and is not acceptable in society. It is so ingrained in human society, that it has been shown that humans unconsciously and automatically apply them to computer agents in social situations[32]. Examples of this include following politeness norms when interacting with a computer agent[33], showing empathy to virtual characters[34] and considering the computer agent as part of a team despite the agent showing no physical human attributes[33]. In the context of building relationships with agents, it has been found that in situations where there errors have occurred, humans are friendlier and more positive to robots who blame themselves rather than their human counterparts, even if the blame should be equally distributed[35].

In multi-agent systems norms are usually set up as a series of permissions and sanctions[36]. Over time, two issues can emerge. Firstly, how to enforce the current norms and secondly how to allow the norms to evolve (to create new norms or modify existing ones) as the system evolves[37]. There have been a number of suggestions for resolving these issues. One proposal is to establish leader and follower behaviour amongst agents to enable new joiners to the system to know who to look to for guidance[38]. Another method is to enable agents to log histories of interactions and identify “special events” which they then confirm with the rest of the system to see if other agents also exhibit the same behaviour[39]. Machine learning, leadership, imitation and data mining have all been used in multi-agent normative systems to study the creation, identification and propagation of norms[40, 41].

3 Constructing a Social Socio-Technical System

Increasingly generalised norms, or values, are being used to assist with designing technology and systems. The concept of value sensitive design[42] is becoming increasingly relevant as society shapes technology and vice versa and the threat of breakdown or malicious attacks of systems increases. One way in which systems breakdowns may be combated is to create a socio-technical system which encompasses key social norms and educates the agents in it to create values which correspond to similar values in human society, thus establishing a “shared reality” between agents.

Before considering how to imbue social values in a socio-technical system, it must first be considered how social values are created in human society. Sociology allows analysis of how human society develops. It also considers how to create a society based on social interactions. One theory in particular which is useful when considering this, is the concept of social constructivism.

3.1 Social Constructivism

The starting point of the proposal is the theory of social constructivism. Social constructivism was initially proposed in the 1960’s by the sociologists Berger and Luckmann in their book “The Social Construction of Reality”[43]. It remains a key work in sociology and has inspired a number of diverging theories in the social sciences[44].

Berger and Luckmann proposed that it is “*knowledge that guides conduct in everyday life*”[43, p33]. One of the main ways in which humans communicate is through language. How we define the objects around us is part of our reality, the very action of talking to one another allows humans to pass on key ideas which then form part of their reality. Our use of language is therefore how knowledge is transferred over time, between generations and cultures. Since we need to communicate to other members in our society to pass on this knowledge, this means that our reality is actually a social construct which we have developed over time. They proposed that reality is composed of two parts, the objective and the subjective. The objective part of reality is the part external to us, the part that as a society we have defined and we all understand to be true. The subjective part of reality is each person’s individual experience of this reality.

To give an example of this, in a global context the term “father” is given to the male parent of a child, every individual reading a newspaper article about a father understands this to be true. The definition of father might have pre-existed them, but as part of their objective reality they nevertheless understand what it means. The subjective reality is that the individual understands that the article is not about *their* father. The link between these two realities is the knowledge that all individuals have the same understanding of the term.

Knowledge is key to being able to function within a socially constructed reality. Individuals need to know current norms of behaviour but also the definitions of objects and ideas which already exist in society. Berger and Luckmann proposed that institutions (or norms) are created through a process called

habitualisation- *“all human activity is subject to habitualisation. Any action that is repeated frequently becomes cast into a pattern, which can then be performed again in the future in the same manner and with the same economical effort”* [43, p70-71]. These repeated patterns of actions introduces institutions, or norms, into society which *“control human conduct by setting up predetermined patterns of conduct”* [43, p72]. According to this theory institutions are not created instantaneously but over time and it is this continual repetition of human actions and interactions which is passed on through language that builds human society.

This concept of habitualisation, learning from patterns over time and then using it to inform future behaviour, is also something which is already extensively used in computing in the form of machine learning.

3.2 Artificial Social Constructivism

Since it is possible to create a shared reality based on social interaction between humans, it is proposed that by imbuing socio-technical systems with human social concepts it is possible to create a shared social reality between the human, computational and digital agents in the network. This could be done by enabling the computational and digital agents access to digitised versions of trust and forgiveness mechanisms seen in human society and by allowing them to learn from each other to establish norms of behaviour which, over time, would develop into values. Although this is more complicated than building up a model of the user and programming it into agents, this method allows the agents to interpret human behaviour more accurately and react accordingly.

Using sociological theories is advantageous in this situation since sociology, which is based on real observed human behaviour, would allow computational and digital agents in the network to better interpret the behaviour of the most irrational agents, humans. This would allow the computational and digital agents to more accurately react and conform to expected human norms. This is particularly important in the occurrence of errors which lead to relationship breakdown.

It is proposed that a computational form of social constructivism, termed artificial social constructivism, is required which encompasses both human, computational and digital agents in a socio-technical system. Artificial social constructivism has three core principles: norms, education and values. The agents will first need to establish norms over time. Then they will need to pass on these established, or changing, norms to both existing and new agents in the system, so that all agents learn how to behave when interacting with one another. Similar to the same way a child brought up in a certain manner comes to view the system’s successes and failures as its own [45] it is posited that through learning, maintaining and enforcing the norms of the system, the agents become invested in it and thus more motivated to uphold it as a whole. In this way the agents will establish key values (generalised norms) which the whole system will adhere to, effectively establishing what in human terms we would call a society.

4 Conclusion

By applying sociological theory to the network in this way, the computational or digital agents should behave as the human agents would expect other humans to behave. The resulting interactions between humans and computers may then be the same as interactions between humans themselves. This has multiple advantages, for example better understanding by agents of other agents' behaviour increases the likelihood of system longevity, but in this position paper we specifically considered the benefit in overcoming relationship breakdowns which occur as a result of errors. As humans automatically respond to computers if they exhibit the same social cues as a human would[33], shared values between agents should lead to a more predictable system from the agents perspectives. The computational and digital agents should be able to better anticipate how the human agents will respond in interactions, while, from the point of view of the human agents, the computer or digital agents will better conform to expected behavioural patterns as part of the human's own social network. The creation of a "shared reality", based on shared experience of social norms and values between the agents in the system, should therefore allow the system to more easily overcome errors by using human traits like trust and forgiveness to repair 'relationships' after faults or misunderstandings between agents occur.

This position paper put forward a review of the relevant literature in trust, forgiveness, norms, values and social capital in socio-technical systems and put forward for a theory which ties these concepts together. Social constructivism introduces a method by which reality is created by human social interactions. Artificial social constructivism, proposed here, is a way of applying this concept to a socio-technical system which uses digital versions of trust, forgiveness, norms, values to create a "shared reality" between humans and computers. The theory aims to address the question of how we can ensure that relationships with, and mediated by, computers are the same as those with other humans. This will in turn allow for the creation of long-lasting socio-technical systems which are better at overcoming errors, and maintaining 'relationships' between the agents, by using already established relationship management mechanisms from sociology.

Acknowledgements

The authors would like to thank the reviewers for their valuable insights and suggestions.

References

1. Nallur, V., Monteil, J., Sammons, T., Bouroche, M., Clarke, S.: Increasing Information in Socio-Technical MAS Considered Contentious. In: 2015 IEEE Int. Conf. Self-Adaptive & Self-Organizing Syst. Workshops, IEEE (Sept. 2015) 25–30

2. Castelfranchi, C., Falcone, R.: Social Trust : A Cognitive Approach. *Trust & Deception in Virtual Societies* (2001) 55–90
3. Hardin, C.D., Higgins, E.T.: Shared Reality: How Social Verification Makes the Subjective Objective. In R. M. Sorrentino, E. T. Higgins, eds.: *Handbook of Motivation and Cognition*. Guilford Press, New York (1996)
4. Ostrom, E., Ahn, T.: *Foundations of Social Capital*. Edward Elgar Pub., Cheltenham (2003)
5. Pitt, J., Nowak, A.: The Reinvention of Social Capital for Socio-Technical Systems. *IEEE Technol. & Soc. Mag.* **33**(1) (2014) 27–34
6. Petruzzi, P.E., Busquets, D., Pitt, J.: Experiments with Social Capital in Multi-Agent Systems. *Lecture Notes in Comput. Sci.* **8861** (2014) 18–33
7. Petruzzi, P.E., Pitt, J., Busquets, D.: Electronic Social Capital for Self-Organising Multi-Agent Systems. *ACM Trans. Autonomous & Adaptive Syst.* **12**(3) (2017) 1–25
8. Gambetta (ed.), D.: *Trust*. Basil Blackwell, Oxford (1990)
9. Luhmann, N.: *Trust and Power*. John Wiley and Sons, Inc, Chichester (1979)
10. Marsh, S.: *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling (1994)
11. Pinyol, I., Sabater-Mir, J.: Computational Trust and Reputation Models for Open Multi-Agent Systems: A Review. *Artificial Intell. Rev.* **40**(1) (2013) 1–25
12. Jones, A.J.I.: On The Concept of Trust. *Decision Support Syst.* **33** (2002) 29–78
13. Cho, J.H., Chan, K., Adali, S.: A Survey on Trust Modeling. *ACM Comput. Surv.* **48**(2) (2015) 40
14. Marsh, S., Dibben, M.R.: Trust, Untrust, Distrust and Mistrust: An Exploration of the Dark(er) Side. *Proc. 3rd Int. Conf., iTrust* (2005) 17–33
15. Mcknight, D.H., Choudhury, V., Kacmar, C.: The Impact of Initial Consumer Trust on Intentions to Transact with a Web Site: A Trust Building Model. *J. Strategic Inform. Syst.* **11** (2002) 297–323
16. Han Yu, Zhiqi Shen, Leung, C., Chunyan Miao, Lesser, V.R.: A Survey of Multi-Agent Trust Management Systems. *IEEE Access* **1** (2013) 35–50
17. Axelrod, R.: An Evolutionary Approach to Norms. *Amer. Political Sci. Rev.* **80**(4) (1986) 1095–1111
18. Pinyol, I., Sabater-Mir, J., Dellunde, P., Paolucci, M.: Reputation-Based Decisions for Logic-Based Cognitive Agents. *Autonomous Agents & Multi-Agent Syst.* **24**(1) (2012) 175–216
19. Rea, T.: Engendering Trust in Electronic Environments. In Castelfranchi, C., Tan, Y.h., eds.: *Trust & Deception in Virtual Societies*. Springer, Dordrecht (2001) 221–236
20. Vasalou, A., Hopfensitz, A., Pitt, J.: In Praise of Forgiveness: Ways for Repairing Trust Breakdowns in One-Off Online Interactions. *Int. J. Human Comput. Stud.* **66**(6) (2008) 466–480
21. McCullough, M.E.: Forgiveness: Who Does It and How Do They Do It? *Current Directions in Psychological Sci.* **10**(6) (2001) 194–197
22. McCullough, M.: *Beyond Revenge: The Evolution of the Forgiveness Instinct*. First edn. Jossey-Bass (2008)
23. Axelrod, R., Wu, J.: How to Cope with Noise in the Iterated Prisoner’s Dilemma. *J. Conflict Resolution* **39**(1) (1995) 183–189
24. Cohen, M.D., Riolo, R.L., Axelrod, R.: The Role of Social Structure in the Maintenance of Cooperative Regimes. *Rationality & Society* **13**(1) (2001) 5–32

25. Vasalou, A., Hopfensitz, A., Pitt, J.: Is an Apology Enough ? How to Resolve Trust Breakdowns in Episodic Online Interactions. In: 21st BCS HCI. Volume 2. (2007) 119–122
26. Lenaerts, T., Martinez-Vaquero, L.A., Han, T.A., Pereira, L.M.: Conditions for the Evolution of Apology and Forgiveness in Populations of Autonomous Agents. AAAI Spring 2016 Symposium on Ethical and Moral (2016) 242–248
27. Ostrom, E.: Collective Action and the Evolution of Social Norms. *J. Economic Perspectives* **14**(3) (2000) 137–158
28. Pitt, J.: Digital Blush: Towards Shame and Embarrassment in Multi-Agent Information Trading Applications. *Cognition, Technol. & Work* **6**(1) (2004) 23–36
29. Mondal, P.: Difference between Norms and Values of Society (2017)
30. Pitt, J.: From Trust and Forgiveness to Social Capital and Justice. In Reif, W., et al., eds.: *Trustworthy Open Self-Organising Systems*. Springer International Publishing, Cham (2016) 185–208
31. Binmore, K.: Social Norms or Social Preferences? *Mind & Society* **9**(2) (2010) 139–157
32. Nass, C., Steuer, J., Tauber, E.R.: Computers are Social Actors. *Conf. Companion on Human Factors in Computing Syst.* (1994) 204
33. Reeves, B., Nass, C.: *The Media Equation*. Cambridge University Press (1996)
34. Paiva, A., et al.: Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. 3rd Int. Joint Conf. Autonomous Agents & Multiagent Syst. (2004) 194–201
35. Groom, V., Chen, J., Johnson, T., Kara, F.A., Nass, C.: Critic, compatriot, or Chump?: Responses to Robot Blame Attribution. In: 2010 5th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), IEEE (2010) 211–218
36. Artikis, A., Sergot, M., Pitt, J.: Specifying Norm-Governed computational Societies. *ACM Trans. Comput. Logic* **10**(1) (2009) 1–42
37. Posner, R.A., Rasmusen, E.B.: Creating and Enforcing Norms, with Special Reference to Sanctions. *Int. Rev. Law and Econ.* **19**(3) (1999) 369–382
38. Savarimuthu, B.T.R., Cranefield, S., Purvis, M., Purvis, M.: Role Model Based Mechanism for Norm Emergence in Artificial Agent Societies. *Lecture Notes in Comput. Sci.* **4870 LNAI** (2008) 203–217
39. Savarimuthu, T., Cranefield, S., Purvis, M.A., Purvis, M.K.: Obligation Norm Identification in Agent Societies. *J. Artificial Societies & Social Simulation* **13**(4) (2010)
40. Savarimuthu, B.T.R., Cranefield, S.: Norm Creation, Spreading and Emergence: A Survey of Simulation Models of Norms in Multi-Agent Systems. *Multiagent & Grid Syst.* **7**(1) (May 2011) 21–54
41. Savarimuthu, B.T.R., Arulanandam, R., Purvis, M.A.: Aspects of Active Norm Learning and the Effect of Lying on Norm Emergence in Agent Societies. In Kinny, D., ed.: *Int. Conf. on Principles & Practice of Multi-Agent Syst.*, Berlin Heidelberg, Springer-Verlag Berlin Heidelberg (2011) 36–50
42. Friedman, B., Kahn, P.: Value Sensitive Design: Theory and Methods. University of Washington Tech. (Dec.) (2002) 1–8
43. Berger, P., Luckmann, T.: *The Social Construction of Reality*. Penguin Books (1966)
44. Knoblauch, H., Wilke, R.: The Common Denominator: The Reception and Impact of Berger and Luckmann’s *The Social Construction of Reality*. *Hum. Stud.* **39** (2016) 51–69
45. Dewey, J.: *Democracy and Education*. Simon & Brown (2011)