The Use of Trust in Social Machines

Arpit Merchant
IIIT, Hyderabad
Telangana, India
arpit.merchant@
research.iiit.ac.in

Tushant Jha
IIIT, Hyderabad
Telangana, India
tushant.jha@
research.iiit.ac.in

Navjyoti Singh
IIIT, Hyderabad
Telangana, India
navjyoti@iiit.ac.in

ABSTRACT

Trust plays an important role in the effective working of Social Machines by allowing for cooperative behaviour amongst human and digital components of the system. A detailed study of trust helps in gaining insights into the working of social machines, and allows designers to create better systems which are able to engage more people and allow for efficient operations. In this paper, we undertake a discussion on the variety of ways in which trust can be observed in Social Machines by outlining a three class taxonomy (personal, social and functional). We build upon earlier observations in past literature while seeking a broader definition. We discuss the problem of trust, that of promoting trust amongst the trustworthy in social machines, and present the various insights, challenges and frontiers that arise in response. This includes the role of institutions, communication processes and value aligned technologies in social, personal and functional trust respectively.

Keywords

Social Machines, Trust, Information Systems

1. INTRODUCTION

What is the problem we wish to solve when we attempt to construct a social machine?

Social Machines are the building blocks of today's networked systems. Internet websites today are no longer simply static pages that one reads. There is active participation from both, the users and the websites. Based on the actions performed by the user, the website software reacts in a specific manner. With recent advances in machine learning, recommender systems, data mining, etc, the response from the software is even intelligent, to a certain extent. For example, if John has already viewed several videos related to (say) Baseball on Youtube, the website can suggest similar videos to him in the future. This evolves over time as changes wrought by past events form patterns in actions and the software adapts to accommodate the new trends. This leap from static to dynamic and intelligent software led to Tim Berners-Lee coining the term Social Machine[3]. A Social Machine is a complex techno-social system comprised of various individuals or groups of individuals and digital components connected through a networked platform, in a predefined mode of interaction for a particular purpose. In recent years, there has been increasing interest in studying Web systems as social machines ([20], [10], [35], [31]).

In our study, we refer to the complete structure of a website along with the users, the content created by the users and the site through interactions with each other, as well as the underlying information processing systems together as a social machine. As an illustrative example, consider Wikipedia, an online, collaborative, free-access encyclopedia created by Jimmy Wales and Larry Sanger[1]. Any person can create, edit, modify and read an article on Wikipedia. A "Talk" page associated with each article allows editors to discuss, coordinate and debate the related issues. In case of disputes, resolution is reached through community consensus and vandalism is restricted through internalized benefits. Editors in good standing can be upgraded to the post of administrators if recommended by the community. This large-scale interaction of humans with machines combines democratic and hierarchical elements to form interleaving social and digital processes. Similarly, Google Plus, Facebook, Github, Twitter, etc. can all be understood through the social machines paradigm. Furthermore, this definition is extensible and can be applied to other systems, as Buregio et al[5] show by constructing Government as a social machine.

The central problem in the construction of a Social Machine (SM) is thus, generation of cooperative behavior between its human and digital components. SMs require complementation of purposes and actions of different actors. And cooperation allows for systematic integration of the merits of the components that constitute it. While cooperative behaviour, as game theory has shown, leads to more effective systems, it might not always be rational under various restrictions due to the distributed nature of the system, possibility of malicious actions etc. In these cases, promoting cooperation requires that there exists trust between the various components. Trust provides the incentives that can ensure mutual guarantees of successful interactions and intended benefits.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890597 .

[1]https://www.wikipedia.org/

The problem therefore, in the context of Social Machines boils down to taking design decisions that foster trust. Trustworthiness[26] is the property of an entity to act in the manner in which it is expected to act whereas trust is the attitude or belief towards an entity that it will act in the expected manner. If trust is placed on an untrustworthy agent or if trust is not placed in a trustworthy agent, then the design of the system is crucially flawed. The solution is thus to get the different components to work together and we posit that by creating a set of incentives and regulations such that trust is placed on that which is trustworthy, better results can be achieved effectively. This is why the study of what trust depends on, how it changes, its nature and complexity is useful in social machines.

To be precise, a unified definition of trust is elusive due to the vast diversity of its constituent facets[23]. The observation and analysis of human behaviour on the web is a hard problem because it involves computational formalization of abstract human notions such as trust. A model that is based on realistic assumptions of trust as a sociological and psychological concept, and can yet be formally specified and rigorously evaluated would lead to a powerful framework for decision making[25]. A complex series of collective reasoning goes into establishing a measure of trust among a group of strangers which is difficult for a machine to grasp. And this might vary from person to person. In order for online systems to exhibit a similar level of response, sophisticated theoretical foundations are required. Krukow et al[19] argue for the need for such a "hybrid" framework of computational trust. In this paper, we define a taxonomy of trust that aims to capture its different elements in the SM paradigm and argue for an extended framework that incorporates those elements that new technological advancements have presented.

The rest of the paper is organized as follows. In Section 2, we present a new taxonomy of trust and describe the three classes. Section 3 presents the problem of trust from the perspective of institutions, information and value alignment problem. Section 4 presents the emergence of a fourth class in the taxonomy of trust, through the context of the Internet of Things. And lastly in Section 5, we present our conclusions and directions for future work.

2. TAXONOMY OF TRUST IN SMs

Trust as a concept, has been studied in a wide variety of disciplines such as philosophy, economics, psychology, sociology and cognitive sciences[23]. Rousseau[28] summarizes contemporary definitions of trust as "...a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions and behaviours of others." Dumouchel[11], on the other hand defines trust as an action. "To trust is to act and not simply to expect because it is the act, not the expectation, that gives the other agent power over the person who trusts." Actions are concrete as they can be observed and their impact can be measured through other actions. However in this context of social machines, exploring the scope of definitions allows us to encapsulate relevant conceptions of trust.

O'Hara[27] discusses some of the challenges of trust in social machines and provides a framework to discuss it. In [27], Trust is defined as a 6-tuple:
X trusts Y $:=_{df}$ = Tr<X, Y, Z, I(R,c), Deg, Warr>
where X and Y are trustor and trustee respectively. Here Z is the agent making claims R about Y's intentions, capacities and behaviours. And I(R,c) is the interpretation of claims R by trustor X in context c. And Trust is seen as the belief held by X that Y will comply by such claims made by Z. Deg denotes the degree of this belief, and Warr denotes the warrant for such belief, rational or irrational. This model, however, is only directly applicable to trust that arises from trustworthiness of systems that comply with some claims and guarantees of the designer. However, as one observes, trust between people and groups of people is also essential to any SM. Such trust is also interesting since, unlike the former, this type of trust usually does not have any Z making claims about the system. Instead, it is deeply rooted in the understanding of individuals about the social context and norms, and mental model they hold of their environment[17].

We present below a threefold model of trust in this technosocial setting based on the locus of trustee in the relationship. We make an obvious distinction between trust that is vested in the software and digital components of system, calling it Functional Trust, from the trust that exists between an individual and his peer in the system, which we call Personal Trust. We further talk about a third class, Social Trust that exists in members of the community with respect to its collective behaviour, which we claim is distinct from both Personal and Functional aspects.

2.1 Personal Trust

This element of Trust in social machines represents the unique features and attributes of the agents themselves. The personality of the individual based on traits such as openness, conscientiousness, neuroticism, extraversion, agreeableness[2] affect how she gives and perceives the receipt of trust. These psychological and other cognitive determinants of trust such as desires, interests of the agent, are valuable characteristics that define their nature[32]. The personal beliefs that individuals hold, colour the way they interact with other individuals. And this trust that one has, is different from peer to peer.

Such trust is vested in one's personality, whose existence is independent of any specific SM. In particular, the trust that the trustor has on the trustee on different SMs is strongly correlated (since there is one common individual behind each of these accounts) and trust on a single platform as such cannot be seen in isolation. Each account is a unique entity in itself with its own history. These accounts not only interact with accounts of others, but also may communicate with each other in a variety of ways for a variety of reasons. For example, one can use one's (say) Google ID to log into (say) Facebook. Yet although each of these SMs are unique in terms of their computational processes, the nature of relationship between humans and machines, their software design and the obtained results, the underlying personality traits behind these each of the actions on the various SMs are common. One's online identity is the sum of one's behaviour across these different SMs.

2.2 Social Trust

Let us say X and Y are cousins and also students studying in the same department. They are both added onto the respective Family and Department groups on Whatsapp. Then the action of trust can be seen to differ considerably

in both the groups despite them being the same people on the same platform. This as we observe, is a consequence of different community roles in which they interpret each other in both cases. This notion of roles in society gives rise to collective behaviour and that impacts how one forms and develops relationships[13]. Moreover, this extends to trusting in the unknown when one does not have enough information to form one's own opinion and instead chooses to trust what the community does.

There is an element of trust comprising of the sociological conditions such as geo-spatial locations of people, the surrounding environment, standard of living, routine habits etc[34]. Culture and traditions that one grows up with influence the way one thinks. One becomes more attuned to those around that belong to similar backgrounds. One identifies more with them and that impacts the manner, speed and nature of trust that develops with them. Additionally, the position of an individual in a society is based on the role(s) that the individual plays[24]. Every society (online or offline) has a structure that governs the way the it operates. A structure might have multiple hierarchies within itself and each hierarchy can have its own set of goals, capabilities and limitations[18]. Apart from these, language and physical characteristics as perceived in the society are also psycho-social determinants of trust.

2.3 Functional Trust

The functional element constitutes of the trust that the individual has on the software and the functionality of the system. Identifying the role and method of functioning of the underlying structure is an important characteristic of how and how much one trusts the system. Perhaps the most important characteristic here is that the claims are rigorously defined while in Social and Individual, there are no binding documents that outline the rules and regulations. One signs "I Agree" on most claims documents without reading them in their entirety. The individuals however, rely their trust not on the detailed analyses of these claims but learn from their experience as in the case of the other elements as well.

This trust that individuals hold in the features of software platform, can be seen in the context of guarantees and standards provided by designer-administrator of platform, and the inherent assumptions such as privacy or absence of malicious code. It is interesting how in many cases the latter, the inherent assumptions, are often not well defined. The difference of interpretation of such norms often leads to conflict and renegotiation of the terms of use of the social machine. This can especially be observed in how the framework of privacy settings has evolved over a period of time as result of various issues and demands of the Social Machine.

Software systems are inherently complex themselves, and appearance and disappearance of software bugs is a regular feature of their engineering cycle. Functional trust also captures the trust in the correctness of programs that constitute various digital components of the SM. This trust can manifest itself in two forms, depending on the organizational style of the software administration, i.e. open source and closed source. In open source, trust for program is redirected to trust in the developer community and *institutions*. On the other hand, in closed source, both the trust on the functionalities as well as their correctness is redirected to trust in the firm(s) that control the digital components.

3. THE PROBLEM OF TRUST

The problem of trust, in the context of design of Social Machines as also discussed in [27], is the promotion of trust amongst human and artificial actors of the Social Machine which are trustworthy, i.e. which will work towards the understood purpose and not behave in a malicious fashion. The hardness of this problem arises from the problems of cooperation and rational behaviour, as established by Hollis in his *Trust within Reason*[16]. The establishment of trust with parties that do not have pre-existing Personal Trust is especially crucial for social machines that involve collaborative participation in some efforts, eg. Wikipedia, Quora, etc. Two common solutions, as also pointed out in [27], include signalling amongst honest agents to establish cooperative behaviour, and establishment of trustworthy institutions that allow honest participation and penalize malicious behaviour. The scope of signalling between honest peers helps them gain more information and establish common context amongst them. Institutions aid in founding of Trust by boiling it down to displacement or redirection of trust from between the peers to towards the community and the procedures and structures established amongst them. Also, while *signalling* and *institutions* help in building trust amongst peers with the aid of technology, many new challenges appear when dealing with trust between humans and complex software systems. With the emergence of intelligent systems and their presence in SMs, for which making claims and guarantees is inherently impossible with state-of-art, new foundations for Functional trust become necessary.

3.1 The Role of Institutions

Institutions play an important role in aligning interests and displacement of trust in social machines, by redirecting trust to hierarchies and procedures established in community. The role of institutions in the functioning of society and economy has been a well-established concept in political science and anthropology[8], including a detailed discussion by Francis Fukuyama in *Trust: The Social Virtues and the Creation of Prosperity*. Fukuyama[14] and subsequent discussions by César Hidalgo[5(u)5], point to the importance of 'trust' in establishment of economic cooperation and distribution and use of information towards shared goals. It has been pointed out how this trust results in emergence of institutions[12], since it is only in the presence of institutions that individuals are able to substitute their trust towards a multitude of peers, many of whom might be unknown to him/her, with a trust towards the institutions through which they interact. Institutions, thus, play a role in simplifying the *act of trusting*.

For instance, while using Wikipedia, despite the absence of personal knowledge about contributors to an article, individuals have a tendency to trust the facts. This is motivated primarily by an *institutional trust* vested on the overall techno-social institution of Wikipedia, which entails its review process as well as a dedicated community. This example in fact, helps us closely analyze the birth of institutional trust, as can be observed in how it essentially involves a combination of social trust in the community as well as functional trust in moderation privileges and other features that support the requisite procedures.

The definition of what constitutes an institution is just as ambiguous[30] in a Social Machine as it is elsewhere. While

at a certain level, a complete social machine can be called an institution, on the other hand, merely the moderation system or the upvote system can also be called an institution in itself. Institutions emerge at various levels in Social Machines. While some institutions might simply be elements of interaction that occur as software features, for instance upvote systems (Quora, Stack Overflow, etc), others might be more complicated hierarchies which are intentionally designed and supported, such as Wikipedia moderation system. However, in some cases they might even emerge unexpectedly in the form of some organizational feature that was not originally intended by the designer. Of course in this case, the designers and administrators usually take notice of such emergent behaviour and have an incentive to evolve the technology front of the social machine to accommodate for them.

3.2 Virtual Spaces and Digital Semiosis

The use of *signalling* or implicit communication of honest intentions has been a well established concept in evolutionary game theory, to combat the negativities of game-theoretic rational behaviour in restrictive settings. Signalling has been argued as one of the foundations of cooperative behaviour in evolutionary biology[7, 6], and is also seen as a act to compensate information asymmetry in economic theory([33], [1]). However, signalling is not an exotic action that requires exquisite planning and effort, but in fact is prevalent in day-to-day human conversation as well, in form of body language and gestures, vocal paralanguage (voice, pitch, rhythm, etc.), and conversational implicature. This, therefore, extends to virtual spaces of communication and interaction created within Social Machines as well, both at textual and symbolic level.

Social Machines that are built on textual communication allow greater linguistic flexibility to establish institutions and trust. An easy way to analyze such communication is to borrow the framework of literary theorist I A Richards, as outlined in [21], that claims four layers of interpretation of text: 1) Surface Meaning or Sense, 2) Emotive Meaning or Feeling, 3) Tone or Attitude, and 4) Intention. Digital communication and conversation can also be seen at these four levels. The Emotive and Attitude layers also emerge as a result of paralinguistic features that emerge in digital communication, such as common practice of interpreting implicit anger and shouting for a text written with all uppercase letters. It can also be observed, then, that trust also plays different roles a different layers. For instance, it is possible for an individual to trust some controversial fact shared publicly by her peer, she might still suspect ill intentions behind it. Consequently, the manner in which each of the layer influences personal trust between peers also differs.

Even in physical conversation, body language and gesture and other forms of non-verbal communication also plays a huge role in enabling signalling. Semiosis, or production of meaning with use of signs and symbols, manifests itself in social machines in form of domain specific non-linguistic structures, such as 'likes', 'pokes', 'shares' on Facebook. Many different features available on different SMs as means of symbolic interactions between peers provide different syntactic systems and established semantics. However, often such symbols can also take other meanings in different personal and social contexts, and consequently help in trust building beyond the intended purposes. For instance, having a large following on Twitter is not an indicator of trust in itself. But it is more likely for someone to believe his tweets to be authentic information if the person has a record of being retweeted massively. And in Social Machines, where communities are expected to interact, the human-computer interaction and symbolic elements of interaction must be designed, taking into cognizance the scope for signalling and trust building.

3.3 Value Alignment of Technology

Functional Trust, as discussed, rests on the trustworthiness of digital components and the liability of the programmer or the designer. However, with a growth of artificially intelligent systems, there is a rise of a new class of complex systems, for whom complete causal reasoning about behaviour is not possible with available knowledge[4]. For simple applications like recommendation engines to slightly more complicated ones like Business Decision Support Systems, these systems, or artificial actors, deal with complex data and are expected to make complex decisions towards their goals. As reasoning about epistemic qualities of these programs, such as about the bias present in presentation of information, becomes increasingly difficult with newer technologies, new challenges appear at the frontier of functional trust.

This problem becomes more severe as the number of operations for which a Social Machine must rely on such artificial actors. And the difficulty of building trust between human components and artificial actors increases proportionally to the amount of complexity of decisions and consequences. This problem has been studied as the *Value Alignment Problem* in Artificial General Intelligence (AGI) and Singularity Studies communities, as the problem of providing bounds on whether the decision system shall not take an action against certain well-understood values[29].

Even weak claims about these systems rely on deep results about the mathematical properties of the structures implemented. This can also be seen as redirection to trust over mathematics as a field of knowledge, which can be seen as an institution and has also been called a Social Machine in itself in a more generalized setting by [22]. However, since precise guarantees about the working of such complex systems cannot be reasoned about, designers and most diligent individuals rely only on broad epistemic and mathematical fundamentals and statistical evidence of performance. Some novel approaches have been taken, such as OpenAI[2], that attempt at generation of trust by opening up the source code and the underlying scientific artifacts to the scrutiny of public. This can be seen as redirection of trust on the community of few diligent participants and their institutions of operation. Nevertheless, it is important to note that most of the trust today for intelligent technologies is still rooted in trust towards expertise of the firm(s) that maintain them.

4. PHYSICAL TRUST AND THE INTERNET OF THINGS

Computers, smart-phones and tablets are no longer the only devices through which one can connect to the Web

[2]https://openai.com/blog/introducing-openai/

today. Internet of Things (IoT) is the network of physical technology that supports the collection and exchange of data and acts on it according to it's instructions[9]. It allows users to connect directly and control remotely the existing infrastructure. A wide range of tools from glasses, watches, electronic devices and appliances even entire homes are now capable of joining a networked platform. With the diversification in instruments of interaction between users, technology and physical reality, these are a growing part of social machines and so we require the trust to be something beyond personal, social and functional. We refer to this as physical trust.

This trust depends on three major factors namely, the security, autonomy and privacy that these devices can provide. Users of these devices assume that they are protected against cyber attacks and cannot be exploited remotely. The innovative software and hardware behind these devices is in its early stages of development and hence is vulnerable to liability. Many of them are incompatible with a host of technologies that are now pervasive. Their capability to be integrated better is crucial to the notion that these devices are indeed more user-friendly and don't end up increasing the workload. These devices also collect data from their users. Since users are limited in terms of their technical knowledge, they trust the manufacturers that this data is being used only for the better functioning of the device itself and not for profiling or mass collection and distribution without explicit and informed consent.

5. CONCLUSION

Trust plays an important role in the way we form relationships, build organizations, interact cooperatively and design responses. However, the definition of trust can vary depending on the person, context, spatio-temporal locations and so on. And therein lies the problem. Trust often acts in ways that are not easily perceivable at different levels of granularity. In scholarly literature, there is no consensus yet on a universally accepted definition of trust. That is because its properties, mechanisms and impact are not fully understood. In this paper, we take a step forward in understanding the role that trust plays in social machines. If a social machine is not trusted, it will not be used and will therefore become useless. However, viewing trust simply in the form of claims made by entities is inadequate. The mental models in the minds of the human users, and the incentives for them to behave the way they say they will are important factors that need to be accounted for. We define trust not just an expectation of positive reactions, but more so as a three-dimensional model that captures the complex nature of interactions that give rise to collective behaviours.

Personal (individual) traits of users, the social collective attitudes of communities and the structural and functional capabilities of the technologies contribute to the formation of trust. We expound on the relationships of these three elements of trust and show how they can result in strong theoretical underpinnings in the design of social machines and the critical analysis of human and societal behaviour. We also show how social machines themselves can be seen as institutions, and how this viewpoint provides an outlook to Internet of Things, virtual spaces and their organizing principles. Further observations about the nature of trust will play a central role in understanding a world with increasing role of technology. Thus this framework is an important step not only for the present, but also for the future of foundational principles of social machines. We leave case studies of trust as it appears in the various forms discussed here for real-world SMs as future work.

6. REFERENCES

[1] G. Akerlof. *The market for 'lemons': Quality uncertainty and the market mechanism*. Springer, 1995.

[2] M. R. Barrick and M. K. Mount. The big five personality dimensions and job performance: A meta-analysis. 1991.

[3] T. Berners-Lee, M. Fischetti, and M. L. Foreword By-Dertouzos. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation, 2000.

[4] N. Bostrom. *Superintelligence: Paths, dangers, strategies*. OUP Oxford, 2014.

[5] V. Burégio, K. Brito, N. Rosa, M. Neto, V. Garcia, and S. Meira. Towards government as a social machine. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 1131–1136. International World Wide Web Conferences Steering Committee, 2015.

[6] S. R. Dall, L.-A. Giraldeau, O. Olsson, J. M. McNamara, and D. W. Stephens. Information and its use by animals in evolutionary ecology. *Trends in ecology & evolution*, 20(4):187–193, 2005.

[7] R. Dawkins and J. R. Krebs. Animal signals: information or manipulation. *Behavioural ecology: An evolutionary approach*, 2:282–309, 1978.

[8] H. Demsetz. Information and efficiency: another viewpoint. *The Journal of Law & Economics*, 12(1):1–22, 1969.

[9] J. Donath. *The social machine: designs for living online*. MIT Press, 2014.

[10] B. R. Duffy. Fundamental issues in affective intelligent social machines. *Open Artificial Intelligence Journal*, 2:21–34, 2008.

[11] P. Dumouchel. Trust as an action. *European journal of Sociology*, 46(03):417–428, 2005.

[12] E. Durkheim. *The rules of sociological method: and selected texts on sociology and its method*. Simon and Schuster, 2014.

[13] M. Foddy, M. J. Platow, and T. Yamagishi. Group-based trust in strangers the role of stereotypes and expectations. *Psychological Science*, 20(4):419–422, 2009.

[14] F. Fukuyama. Trust: The social virtues and the creation of prosperity. *World and I*, 10:264–268, 1995.

[15] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[16] M. Hollis. *Trust within reason*. Cambridge University Press, 1998.

[17] D. E. Kieras and S. Bovair. The role of a mental model in learning to operate a device*. *Cognitive science*, 8(3):255–273, 1984.

[18] R. M. Kramer. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual review of psychology*, 50(1):569–598, 1999.

[19] K. Krukow, M. Nielsen, and V. Sassone. Probabilistic computational trust. 2009.

[20] M. Luczak-Roesch, R. Tinati, and N. Shadbolt. When resources collide: Towards a theory of coincidence in information spaces. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 1137–1142. International World Wide Web Conferences Steering Committee, 2015.

[21] B. Malinowski, F. G. Crookshank, C. K. Ogden, and I. A. Richards. *The Meaning of Meaning; a Study of the Influence of Language Upon Thought and of the Science of Symbolism*. Kegan Paul, Trench, Trubner; Harcourt, Brace, 1923.

[22] U. Martin and A. Pease. Mathematical practice, crowdsourcing, and social machines. In *Intelligent Computer Mathematics*, pages 98–119. Springer, 2013.

[23] D. H. McKnight and N. L. Chervany. The meanings of trust. 1996.

[24] D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt, and J. Stewart. Social collective intelligence. 2014.

[25] L. Mui. *Computational models of trust and reputation: Agents, evolutionary games, and social networks*. PhD thesis, Massachusetts Institute of Technology, 2002.

[26] K. O'Hara. A general definition of trust. 2009.

[27] K. O'Hara. Trust in social machines: The challenges. 2012.

[28] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer. Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404, 1998.

[29] S. Russell, D. Dewey, M. Tegmark, J. Kramar, and R. Mallah. Research priorities for robust and beneficial artificial intelligence. *Future of Life Institute*, 2015.

[30] J. Searle. *Making the social world: The structure of human civilization*. Oxford University Press, 2009.

[31] N. R. Shadbolt, D. A. Smith, E. Simperl, M. Van Kleek, Y. Yang, and W. Hall. Towards a classification framework for social machines. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 905–912. International World Wide Web Conferences Steering Committee, 2013.

[32] W. Sherchan, S. Nepal, and C. Paris. A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, 45(4):47, 2013.

[33] M. Spence. Job market signaling. *The quarterly journal of Economics*, pages 355–374, 1973.

[34] M. Tanis and T. Postmes. A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, 35(3):413–424, 2005.

[35] M. Van Kleek, D. A. Smith, D. Murray-Rust, A. Guy, K. O'Hara, L. Dragan, and N. R. Shadbolt. Social personal data stores: the nuclei of decentralised social machines. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 1155–1160. International World Wide Web Conferences Steering Committee, 2015.