Lecture Notes in Social Networks

Nigel Shadbolt
Kieron O'Hara
David De Roure
Wendy Hall

The Theory and Practice of Social Machines

Springer

Lecture Notes in Social Networks

More information about this series at http://www.springer.com/series/8768

Nigel Shadbolt • Kieron O'Hara
David De Roure • Wendy Hall

The Theory and Practice of Social Machines

Springer

Nigel Shadbolt
Department of Computer Science
University of Oxford
Oxford, UK

Kieron O'Hara
Electronics and Computer Science
University of Southampton
Southampton, UK

David De Roure
Oxford eResearch Centre
University of Oxford
Oxford, UK

Wendy Hall
Electronics and Computer Science
University of Southampton
Southampton, UK

ISSN 2190-5428 ISSN 2190-5436 (electronic)
Lecture Notes in Social Networks
ISBN 978-3-030-10888-5 ISBN 978-3-030-10889-2 (eBook)
https://doi.org/10.1007/978-3-030-10889-2

Library of Congress Control Number: 2018968363

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface and Acknowledgments

Social machines are a type of social interaction or enterprise that is, in one sense, as old as the hills, and in another, as novel as the technologies that underlie them—the World Wide Web, social networking, smartphones, and so on. We believe, in accordance with the interdisciplinary study of these technologies that has been called *Web Science* (Berners-Lee et al. 2006), that they deserve a place in our standard vocabulary of social cooperation and interaction, thanks in large part to the transformative effects of the increased penetration of society by technology, and to the feedback loops that result as that massive social uptake transforms the technology in return. We therefore focus in this book on the novelty of social machines, although we try not to forget that they are rooted in time-honoured forms of interaction.

This new, if not unprecedented, sociotechnical phenomenon has been the object of study of a programme grant awarded by the UK's Engineering and Physical Sciences Research Council (EPSRC), entitled SOCIAM: The Theory and Practice of Social Machines, and this book describes the work of the Sociam project. The authors have all worked on, or been associated with, Sociam, from the award of the funding in 2012 until the close of the project in 2018, at its four partners: Electronics and Computer Science, University of Southampton; Dept. of Computer Science, University of Oxford; the Oxford e-Research Centre; and the Dept. of Informatics, University of Edinburgh. Our aim has been to map the space of social machines, to understand their development, and to produce an inventory of the skills required to analyse and engineer them and make them sustainable and effective. This is clearly a multidisciplinary imperative, and a complete picture of social machines will draw upon social science, law and regulation, network science, data science and computer science (Tiropanis et al. 2015). Our focus here is on the engineering and data elements of this coalition, but we should always remember that social machines are social as well as machines. They cannot be built; technology may suggest a certain type of interaction and provide a palette of communications, but without participants they are dead formalisms.

The authors were but a small part of Sociam. Shadbolt was its Principal Investigator, De Roure and Hall Co-Investigators, and O'Hara a postdoctoral researcher on the project. The authors are responsible for this book, and any mistakes

or errors contained within it, but the narrative draws on the work of dozens more researchers across the four sites over a six-year period. This work is referenced, so the originators of the ideas, experiments and studies should be clear; the sheer range and fertility of the research should make it abundantly clear that we four authors are reporting on a collective enterprise which would not have been possible without the far-sighted funding model of the EPRSC. We would like to acknowledge these researchers in this Preface.

To begin with, our thanks for the contributions and direction of the other Co-Investigators, Peter Buneman of the University of Edinburgh, Luc Moreau of the University of Southampton (now at Kings College, London), Dave Robertson of the University of Edinburgh, and Tim Berners-Lee of M.I.T. and the University of Southampton (now at the University of Oxford), the last of whom, as we shall see, originated the social machine concept in its modern form. Vital strategic input was provided by our critical friends on our international advisory board, which was chaired by Prof. Lynda Hardman of CWI and the University of Utrecht. The other board members were Prof. Avi Bernstein of the University of Zürich, Prof. Jon Crowcroft of the University of Cambridge, Prof. Juliana Freire of New York University, Prof. Fausto Giunchiglia of the University of Trento and Prof. Chris Welty of Google Research and the VU Amsterdam. Thanks as well to our EPSRC representatives on the advisory board, Lisa Coles, Ellie Gilvin and Richard Gunn, and particular thanks to our project administrator Susan Davies.

Faculty members associated with Sociam included Claudia Pagliari of Edinburgh, Chris Lintott and Ursula Martin of Oxford, and Jacek Brodzki, Les Carr, Adriane Chapman, Hugh Glaser, Elena Simperl, Steffen Staab and Thanassis Tiropanis of Southampton. Core postdoctoral researchers funded directly by Sociam included Alan Davoust, Adrià Gascón, Egor Kostylev, Areti Manataki, Dave Murray-Rust and Petros Papapanagiotou of Edinburgh, Reuben Binns, Megan Meredith-Lobay, Grant Miller, Kevin Page, John Pybus, Rob Simpson, Ségolène Tarte, Milo Ian Thurston, Max Van Kleek and Jun Zhao of Oxford, and Gianluca Correndo, Dominic DiFranzo, Laura Drăgan, Trung Dong Huynh, Markus Luczak-Roesch, Heather Packer, Paul Smart, Daniel Smith and Ramine Tinati of Southampton. Postdoctoral researchers who collaborated with the project included Joe Corneli, Jason Nurse, Petar Radanliev, Nate Rosenblatt and Pip Willcox of Oxford, and Aastha Madaan, Conor Smyth and Xin Wang of Southampton. Visiting researchers included Davide Ceolin at Southampton. PhD students funded directly by Sociam included Tom Cairns, Gianfranco Cecconi, Frode Hegland, Michał Hoffman, Vanessa Mathews, Philip Sheldrake and Henry Story of Southampton. PhD students who collaborated with the project included Amy Guy of Edinburgh, Ulrik Lyngs and Gatis Mikelsons of Oxford, and Ian Brown, Seyi Feyisetan, Darren Richardson, Peter West, Yang Yang and Michael Yip of Southampton. Student interns, who carried out important implementation work, included Keir Bowater, Andrew Francis, Sauyon Lee, Dean Ottewell, Michaela Peykovska and Adam Slack of Oxford, and Gustavo Buzogany Eboli, Yuki Ikuno, Daniela Mihai, Neal Reeves, Filip Stoyanov and Anna Weston of Southampton. The book you hold in your hands, or view on your screen, could not have been written without this excellent team.

We have on occasion reviewed unpublished work by fellow Sociam members, which makes it hard to acknowledge, except explicitly in a footnote. Peter Buneman's ideas on communication and annotation, and also (in collaboration with Adriane Chapman) on social knowledge machines come under this category, as does work by Dave Robertson on expressing the properties of social machines (reported in the section on ethical functionalism in Chap. 4), and Jacek Brodzki's work on topological data analysis reported in Chap. 5 as an important future direction for social machine research. Also, some of Sociam's supporters and collaborators provided work that 'filled in some gaps' in our account, and upon which we have therefore drawn; this covers contributions by Joe Corneli of the University of Oxford, and Richard Giordano and Jeff Vass of the University of Southampton, whom we would like to thank and acknowledge. To reiterate, any errors that have crept into our reports of their research are our responsibility alone.

Finally, thanks to the copyright holders of several images for permission to reproduce them; the papers from which they have been taken are given in the figure captions. Figures 2.1, 2.4, 2.5, 2.6, 3.8, and 3.9 reprinted with permission from Springer Nature. Figures 2.7, 2.8, and 2.9 copyright © 2015 Dave Murray-Rust et al., licensed to EAI, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/3.0/). Figure 2.10 copyright © 2013 World Wide Web Consortium, (MIT, ERCIM, Keio, Beihang), http://www.w3.org/Consortium/Legal/2015/doc-license. Figure 2.11 copyright © 2016 Luc Moreau, reprinted with permission. Figure 2.14 copyright © 2014 Laura Drăgan, Markus Luczak-Roesch, Elena Simperl, Bettina Berendt & Luc Moreau, reprinted with permission. Figure 2.15 copyright © 2015 Laura Drăgan, Markus Luczak-Rösch, Bettina Berendt, Elena Simperl, Heather Packer and Luc Moreau, reprinted with permission. Figure 3.1 reprinted with permission from Mary Ann Liebert, Inc., New Rochelle, NY. Figure 3.2 © 2013 IEEE, reprinted, with permission, from *IEEE Intelligent Systems*. Figure 3.3 © 2017 Oluwaseyi Feyisetan, reprinted with permission. Figures 3.4, 3.5, 3.6, and 3.7 reprinted by permission from Zooniverse. Figure 4.7 copyright © 2016 Claudia Pagliari and Santosh Vijaykumar, under the terms of the Creative Commons Attribution License. Figures 4.9 and 4.10 reprinted with permission of SCITEPRESS, image created by Vanessa Handschke. Figure 5.3 copyright © 2016 Jacek Brodzki and Conor Smyth, image created using the Ayasdi platform, reprinted with permission.

Oxford, UK	Nigel Shadbolt
Southampton, UK	Kieron O'Hara
Oxford, UK	David De Roure
Southampton, UK	Wendy Hall

Contents

About the Authors

Nigel Shadbolt FRS FREng is the Principal of Jesus College, Oxford, and a professorial research fellow in Computer Science at the University of Oxford, UK. He is the chairman and co-founder of the Open Data Institute, and was an advisor on open data to the UK governments of both Gordon Brown and David Cameron, appointed to the Public Sector Transparency Board in 2010. He helped lead the creation of government open data website data.gov.uk. His research interests are the formation and emergence of intelligent behaviour in humans, machines and systems such as the World Wide Web, and is an interdisciplinary researcher ranging over psychology, artificial intelligence, computer science, cognitive science and Web science. He was elected a Fellow of the Royal Society in 2017, and was President of the British Computer Society in its 50th anniversary year. He is the author of *The Spy in the Coffee Machine*(with Kieron O'Hara) and *The Digital Ape* (with Roger Hampson). He was the Principal Investigator on the Sociam programme grant.

Kieron O'Hara is an associate professor in Electronics and Computer Science at the University of Southampton, UK. His research interests are around the nature of digital modernity: the impact and future trajectory of networked digital technology on society, the economy and politics, with a focus on trust and privacy, Web Science, data science and the World Wide Web. He is the author of several books, including *Trust From Socrates to Spin*, *The Spy in the Coffee Machine* (with Nigel Shadbolt), *Conservatism*, and *The Anonymisation Decision-Making Network* (with Mark Elliot et al.), as well as book-length studies of Joseph Conrad, Aldous Huxley, and the 1975 British European Community Referendum. He is one of the leads of the UKAN network of data anonymisation professionals.

David De Roure is professor of e-Research at the University of Oxford, UK, and was Director of the Oxford e-Research Centre from 2012 to 2017. He has strategic responsibility for Digital Humanities at Oxford within The Oxford Research Centre in the Humanities (TORCH). Focused on advancing digital scholarship, he works across multiple disciplines including social sciences, computational musicology, experimental humanities, the Internet of Things, large scale distributed systems and social computing. He has extensive experience in hypertext, Web Science and Linked Data. He was closely involved in the UK e-Science programme and from 2009 to 2013 was the UK National Strategic Director for Digital Social Research for the UK Economic and Social Research Council, and subsequently Strategic Advisor for new and emerging forms of data and real time analytics.

Dame Wendy Hall FRS FREng is Regius Professor of Computer Science at the University of Southampton, UK, and is an Executive Director of the Web Science Institute at Southampton. She was Dean of the Faculty of Physical Science and Engineering from 2010 to 2014, and was Head of the School of Electronics and Computer Science from 2002 to 2007. One of the first computer scientists to undertake serious research in multimedia and hypermedia, she has been at its forefront ever since. The influence of her work has been significant in many areas including digital libraries, the development of the Semantic Web, and the emerging research discipline of Web Science. She was elected a Fellow of the Royal Society in 2009, President of the British Computer Society (2003–2004).and President of the Association for Computing Machinery (ACM) in 2008, the first person from outside North America to hold this position. She was co-Chair of the UK government's AI Review in 2017, and became its first Skills Champion for AI in the UK in 2018.

Chapter 1
Characterising Social Machines

Introduction

In 2007 and into 2008, a disputed presidential election in Kenya ended in a tragic and complex pattern of violent protests, heavy-handed responses from the police and ethnically targeted massacres. With conflict between tribes, between urban and rural and between rich and poor, making sense of the unfolding chaos was nearly impossible. Yet Kenya is relatively well connected by digital technologies, allowing the power of information and the tools to manage it to be brought to bear. A blogger proposed that eyewitness reports and photographs of the violence might be uploaded to a website, and placed on a Google Map (Okolloh 2009). A website—called *Ushahidi*, a Swahili word meaning 'witness' or 'testimony'—was created within days, allowing the cooperative creation of a dynamic map of the violence, with contributions by email and text. The widespread penetration of mobile technology allowed the creation of a virtual community of witnesses documenting violence and human rights violations together. This crowdsourced resource complemented mainstream media reports as well as the blogs of citizen journalists, filling in a number of gaps where some violence was not reported anywhere else, evolving in real time, with a greater specificity of location, and extending over a wider geographical area (Meier and Brodock 2008). The Ushahidi software was later repackaged, to be reused in new emergency circumstances where information is lacking and time is short, such as the earthquake in Port-au-Prince, Haiti, in 2010, where a new map of the devastated city was created for real-time situational awareness (Morrow et al. 2011). Crowdsourcing maps have now become a standard way of creating maps for crises; for instance, Amnesty International has crowdsourced maps to try to trace human rights abuses in Darfur (https://decoders.amnesty.org/projects/decode-darfur), while several projects created more accurate maps than those already existing during a number of outbreaks of Ebola in Africa from 2014 onwards (Lüge 2015).

N. Shadbolt et al., *The Theory and Practice of Social Machines*, Lecture Notes in Social Networks, https://doi.org/10.1007/978-3-030-10889-2_1

Ushahidi and its offshoots are an example of a phenomenon made possible by technologies such as the Internet, the smartphone, social media and the read-write World Wide Web (Berners-Lee and O'Hara 2013), connecting people at scale to document situations, cooperate on tasks, exchange information or even simply play. The key here is cooperation and organisation; social change is not effected just by the spread of information through a network. A group or enterprise must form, able to reproduce itself, and for that, information, incentives and wider culture must align (Cebrian et al. 2016). The interactions involved might be coordinated by some central organising principle in the software, even while participants repurpose the functionality they find to pursue their own projects and agendas. Existing social processes may be scaled up, and new social processes enabled, to solve problems, augment reality, create new sources of value and disrupt existing practice. We call these phenomena *social machines*, built on the foundation of *social computing* (O'Hara et al. 2013).

The power of social computing can be dramatic, as for example with the response to the DARPA challenge of 2009 (Robertson and Giunchiglia 2013). In this artificial context, the aim was to find ten weather balloons placed randomly around the United States (in nine different states from California to Delaware). The rules of the challenge were intended to support the growth of a network of people taking part in the search, enabling a crowdsourced solution. The means of doing this in the winning solution (from Sandy Pentland at the Massachusetts Institute of Technology) was to set out financial incentives according to a Query Incentive Network Model (Kleinberg and Raghavan 2005), in which people were incentivised both to look for the balloons and to add more people to the network. People were paid a proportion of the prize money, depending on whether they found a balloon, or—crucially—whether they introduced someone to the network who also received prize money. This meant that the network members had incentives to cooperate with each other, and to recruit new members of the team, so that finding the balloons was not a zero sum game. Pentland's team began with 4 people, and using social media had recruited over 5000 at the point of completion of the task, which extraordinarily took under 10 h (Pickard et al. 2011).

In the intervening period since these initial experiments, social computing and groups of networked individuals connected by technology have evolved and grown into a much more common phenomenon. We will give examples throughout this book, and introduce some in more detail later in this chapter (as we will see, the social machine world is highly mutable, and so many of these examples will have come and gone in time—they serve to illustrate). Computer-mediated interactions have enabled communities to provide a social response to modern problems of transport, such as Waze, https://www.waze.com/, a navigation app which uses community-derived real-time data about incidents such as traffic jams and accidents, and problems of crime, such as BlueServo, http://www.blueservo.net/, which crowdsources policing on the Texas-Mexico border, and Onde Tem Tiroteio (Where the Shootouts Are), https://www.ondetemtiroteio.com.br/, which uses a network of a million people on Facebook, Twitter, Instagram, WhatsApp, Telegram and a special-purpose app to provide real-time information about shootings and gang-related crime in Brazil. Crowdsourced journalism has focused around the hashtag #DigitalSherlocks, used for example by two BBC journalists to authenticate and geolocate video of an atrocity committed (and denied) by the army in Cameroon (https://www.youtube.com/watch?v=XbnLkc6r3yc). There is a growing number of social machines in the area of health (Van Kleek et al. 2013) to enable those suffering from a particular healthcare problem to pool resources and to offer support and advice to fellow sufferers (see below, and Chap. 4). Social machines spring up to coordinate disaster response from the ground up, such as the Cajun Navy, in which leisure boat owners in Louisiana located victims of floods in 2016 via a Facebook group (https://www.facebook.com/groups/TheCajunNavy/), and coordinated rescues using GPS app Glympse (https://www.glympse.com/) and walkie-talkie app Zello (https://zello.com/). The game of freeform or freestyle chess, online chess where players are allowed to draw upon unlimited alternative resources including other human players and chess computers, is being played by increasingly large groups of players which could be described as social machines, so games take place between two social machines (Hendler and Mulvehill 2016, 47). Some social machines help volunteers to contribute to labour-intensive scientific projects ('citizen science'), such as Zooniverse (see below, and Chap. 3); Foldit (http://fold.it/portal/), an online game about protein folding; myExperiment (https://www.myexperiment.org/), a social website for sharing scientific workflows; iSpot (https://www.ispotnature.org/communities/uk-and-ireland), a social network for nature lovers; and Climateprediction.net (http://www.climateprediction.net/), a distributed computing project to improve climate modelling. edX (https://www.edx.org/) is a MOOC provider that uses its data to research into education. Stack Overflow (https://stackoverflow.com/) is a social network providing programming advice. Crowd Map the Crusades (http://dhcrowdscribe.com/crowdmap-the-crusades/) is a project to crowdsource maps and other visualisations relating to a twelfth-century poem about the First Crusade. These social machines often use collaborative resources such as Facebook, Twitter, GitHub or HUBzero.

It will be obvious from examples such as BlueServo that such efforts will not always be uncontroversial. Attempts to crowdsource the identities of the bombers of the Boston Marathon in 2013 bordered on farce, and, although the countercultural website 4chan (http://www.4chan.org/) was prominent in the home-made policing efforts with its so-called 4chan Think Tank, its lamentable attempts were soon parodied elsewhere on the same site (Walker 2013; Nhan et al. 2017). Twitter provides a platform that facilitates increased interest in events like the Boston bombing, and local news and law enforcement agencies become important hubs on that platform (Buntain et al. 2016); such broadcasting models are less like social machines, because the ambition of social interaction is limited.

The term 'social machine' has been around for a while, without generating a consensus about what it might mean. It has been used pejoratively, as for example by the novelist Norman Mailer (1968), and with a deliberate air of paradox, for instance by Deleuze and Guattari (1977). The spirit of the interpretation we put forward here is captured by Berners-Lee in his book *Weaving the Web*.

> Real life is and must be full of all kinds of social constraint—the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which people do the creative work and the machine does the administration. (Berners-Lee and Fischetti 1999, 172, authors' emphasis)

To be sure, this is an optimistic view. Like all tools, they may be used for good or ill, and we will exhibit some antisocial machines to prove it. But ultimately, social machines have the potential to empower, via a democratic spirit of cooperation and respect. They bring people together, using technology to enable communication and cooperation at scale.

The concept does seem to be one which reaches beyond the research community. A class on social machines has been taught at the Digital Humanities Summer School at the University of Oxford since 2015, and has shown that students there find little difficulty with the concept, and are able to design their own social machines, thinking fairly naturally in terms of the roles, interactions, incentives, information requirements and value propositions needed to structure the machines and make them sustainable. Students working in groups have been able to think about the design of social machines to curate and participate in exhibiting material from a museum's collection; crowdsource advice about where vegans can eat out in Japan; contribute private material to a digital archive; use Zen techniques to mitigate anxiety and stress; and crowdsource and verify traffic reports. Hence social machines and the ideas behind them, we would argue, are not alien concepts to people who are digitally literate.

In Fig. 1.1, we see the dimensions along which growth has taken place in order to make social machines possible. In the first place, the complexity of the computation (both in terms of software and hardware) and the quantity of the data

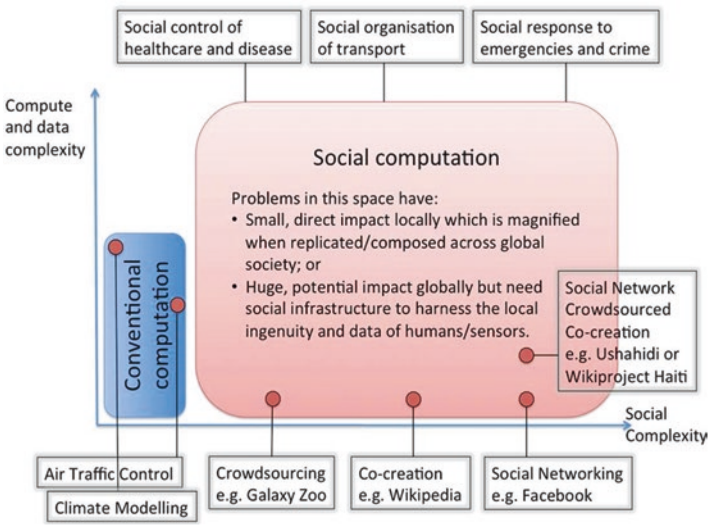

Fig. 1.1 Increasing complexity in computation and social interaction

increase, represented here along the vertical axis. Because of this increase, highly difficult safety-critical and data-heavy problems can now be addressed, from air traffic control to modelling patterns of climate. Deep learning is becoming increasingly effective in real time, and data-driven artificial intelligence is now a routine research tool. Along the horizontal axis, we chart increases in social complexity as technology allows dispersed communities to form and cooperate in order to solve problems, such as labour-intensive scientific research (we will see the example of Galaxy Zoo later in the chapter and in Chap. 3), co-creation of content, as in Wikipedia (also discussed below, and in Chap. 3), or social networking. Ushahidi, with its more complex computing requirements, begins to push the frontier to the upper right. If we manage to develop tools and techniques to populate the upper right of the graph, where computation and social interaction are both more complex than current systems can support, then one could imagine the formation of social machines to manage difficult problems in healthcare, transport or policing.

Social machines are co-created by their human participants and their technological components. The use of the term ‘co-creation’ is deliberate—there is no simple dividing line between the human and the technological, and from the point of view of the social machine it may be irrelevant whether a particular function is performed by a person or a computer (as we shall see, it is impossible to stick to Berners-Lee’s condition of reserving the administration for the computer and the creative work for the people). This first chapter will attempt to make the idea of social machines a little more concrete, but we can say immediately that technology and practice allow certain characteristics to emerge:

- Social machines can work at large scales (numbers of participants), harnessing the power of the crowd.
- They can work in real time, as their participants cooperate in solving a problem.
- They can exhibit network effects, where the value to the participants increases exponentially as the number of participants increases.
- Thanks to the properties of the relevant digital technologies, they can generate, and have access to, large quantities of data.
- All participants can make significant contributions. As with the example of Ushahidi, no-one has the whole picture. *No-one knows everything, but everyone knows something.* If those somethings can be aggregated effectively, the result is a wider picture, hopefully more reflective of participants’ interests, than could have been produced by a small group of experts.

In this book, we explore the possibilities and new realities that these characteristics afford. In the remainder of this introductory chapter, we refine our concept of ‘social machine’, with some examples, classifications and connections with other fields of study.

What Is a Social Machine? The Machine Metaphor

It is important to understand at the outset that the term 'social machine' is a metaphor. It is certainly not meant to imply that social processes can all be made mechanical (though they may have mechanistic properties), and particularly not that their mechanistic properties imply that the processes are determined in advance, physically necessary or radically limited by the technology. We also do not follow Hendler and Mulvehill in including AI machines able 'to interact in ways that have traditionally been considered the social space of humans only' (2016, 10). Such machines aren't socially *constituted*, though they may be able to negotiate social situations. Neither is a social machine a literal *machine*. Instead, a social process, joint endeavour or enterprise is *likened* to a machine in order to highlight certain of its aspects and render them available for study.

Many will resist the metaphor or give it a pejorative force; Mailer (1968) excoriated the guests at a *bien pensant* cocktail party as the 'servants of that social machine of the future in which all irrational human conflict would be resolved, all conflict of interest negotiated, and nature's resonance condensed into frequencies that could comfortably phase nature in or out as you please'. Nevertheless, it is intriguing to note that the word 'machine' used routinely to be applied to social phenomena in English, where until the seventeenth century it might refer to a plot (we still use the related word 'machination'), whose machine-like aspects included the application of power and the mechanical unfolding of the plan without the conscious cooperation of the 'cogs' and 'components'. Queen Elizabeth I, in a letter to James VI of Scotland, hoped that 'our solide amitie may overthawrt thes develische machines'. American political parties, especially in big cities, used to be accused of indulging in 'machine politics', i.e. working to get the vote out reliably by providing services and handouts close to election time, rather than debating about ideology or trying to convert opponents. Even now the machine metaphor is recognisable, as in Nick Cave's song *Red Right Hand*, which warns the listener 'you're one microscopic cog in his catastrophic plan/Designed and directed by his red right hand'.

This sort of scheme or plot is not, however, the meaning we wish to convey here. Social machines are *social* machines, in which people are not components, but *participants*. A fascinating early use of the term from 1846 is Captain William Allen's *Mutual Improvement: or a Scheme for the Self-Adjustment of the Social Machine*,[1] in which 'the mighty engine of PUBLIC OPINION' is put to use to promote virtue and suppress vice. Capt. Allen espouses the key social machine intuition, which is that no-one knows everything, but everyone knows something: 'while the power of selecting the most deserving is rendered difficult to the distributors of awards by a variety of obstacles, such as prejudices, false appearances, the impossibility of appreciating real motives, and of tracking out secret practices, which can belong only to omniscience and omnipresence; this power is possessed, in a great degree,

[1] Thanks to Ian Brown for discovering this reference.

by every member of a community, who is at the same time the censor and under the censorship of his comrades collectively and individually'. The mighty engine is deployed in specific human communities (Capt. Allen proposes to apply his method: '(1) To seminaries for the instruction of youth of all classes. (2) To communities of the lower orders of manufacturing and rural districts. (3) To military professions. (4) To prison discipline'.), all members of which are expected to participate in the promotion of virtue. Throughout this book, we therefore use the term 'participants' to refer to humans in the social machine (there are perspectives, such as Actor-Network Theory (ANT—Latour 2005) that refuse to make a distinction between humans and technologies in cooperative ventures, but we do not adopt that view). The bland term 'user' is not appropriate, because in a social machine the participants drive the interactions; the social machine evolves with the aspirations of its participants, so it would be a mistake to suggest that they simply *use* a given technology. The constitutive relationship, as argued, is co-creation.

Finally, it is important to note that the technology and the social practices around it are changing rapidly—sometimes it is hard to remember that neither social networking nor smartphones were widespread in the mid-2000s. Opportunities to use technology to achieve ends have multiplied rapidly in that time, and entrepreneurial people and communities have developed fascinating and unanticipated applications for the digital technologies to hand.

Ten Aspects of Machines

Many interesting aspects of machines contribute to the image of a social machine. To fill out the metaphor, consider the following ten. The last two are properties of a specific type of machine, the computing machine; the preceding eight are perfectly general.

1. A machine uses energy to do **work**. A social machine uses the 'energy' created in human interaction to achieve something by channelling that energy and changing (or preserving, or reproducing) some aspect of the world.
2. A machine exploits **mechanical advantage**, that is, the amplification of input force. A social machine amplifies the energy of human interaction by exploiting the scale made possible by the Web in real time.
3. A machine has a **material** realisation in some human artefact. The Web and other networked technologies are essential to realise a social machine.
4. A machine (usually) has moving **parts**, and can be disassembled. The 'parts' of a social machine are the technologies and the participants. The contrast is that these 'parts' are not under central control, and a participant of a social machine can generally walk away.
5. A machine is **coherent**. It is an assembly that is independent enough of its environment to make us wish to name it, to refer to it in isolation (even though it may not be able to function in isolation). Furthermore, we name it in virtue of

its coherent function—if we disassembled the machine and laid its parts out on the floor, or merely identified the parts needed to build the machine from an inventory of stock at B&Q or Home Depot, then the imperative to name and refer would be absent.

6. A machine is **designed**. The technology within a social machine is designed to constrain the human interaction to amplify its effects (although the participants can always engender new interactions via some alternative channel, such as a social networking site, so the design of the social machine is not a given, but a dialectic).
7. A machine usually has a ***telos***, a general purpose or a more specific goal. This does not imply any anticipated result; indeed it may simply be of aesthetic value, as for example with the fantastical designs of the cartoonist W. Heath Robinson. The participants in a social machine may take part simply because they enjoy the interaction, not because they are concerned with the purpose of the social machine (or it may not have a purpose beyond acting as a setting for its participants' interactions). The *telos* can sometimes be read off from the design.
8. The perspective of the machine is **emergent** from the actions of the components. The intentions of the participants in a social machine need not coincide with the *telos* of the machine as a whole, and indeed the participants might repurpose the machine by their own participation.
9. A computer is a special type of machine that **processes information** in a programmed sequence. A social machine can also be seen as processing information in sequence, although the 'programme' may be emergent from the operation of the machine. Processing might be done by the architecture harnessing the contributions of a crowd, by individual participants or by participants working together.
10. A computer holds information in **memory**. The participants bring information to a social machine, which can be exploited in the processing.

Any particular social machine is likely to have some, probably most, and occasionally all of the digital analogues of these properties. These are not to be taken as necessary or sufficient conditions of social machinehood—rather, if a social use of digital technology can be understood in a significant proportion of these ways, then we will find ourselves tempted to use the term 'social machine' to describe it.

Mechanistic Explanation

The use of the machine metaphor implies that within a social machine the processes are understood as 'mechanistic', that is, the explanation of the processes is, or is in large part, mechanistic. An explanation of a process is mechanistic if it is couched in terms of the causal structure of the process, i.e. the ways in which the constituents (including human participants) interact with each other and causally influence behaviour. A mechanistic explanation implies (1) some phenomenon/behaviour to be explained (which may have been intended by a designer, but need not necessarily have been); (2) a set of components whose interactions constitute the system at least in part ('in part' because the system could also rely on other elements from the environment, as for example an electricity supply that is not strictly part of the machine, or an external load); and (3) organisation or a structure of these components, constraining the potential interactions (Salmon 1998; Craver 2007).

Note that we should not read too much into a mechanistic explanation of social machines' behaviour.

- The existence of a mechanistic explanation does not preclude other types of explanation of the system.
- It neither implies that the behaviour to be explained was physically necessary (i.e. that it could not have been otherwise given the existence of the mechanisms in the explanans), nor suggests that social machines' behaviour could be predicted from the structure of their participants and technological constituent parts.
- It does not assume that the machine (or its design) is in any way prior to the behaviour of the constituents: It may be (and often is) that the participants' interactions enable a social machine to become detectable (or to 'emerge'), and that the participants may change their behaviour in such a way as to change the emergent social machine.
- Finally, it does not mean that the behaviour of a human participant in the system can itself be mechanistically explained (i.e. the social machine paradigm does *not* reduce the behaviour of participants to mechanisms, except *qua* constituent of the social machine). Whether human behaviour can never, sometimes or always be explained mechanistically is neither here nor there; as far as the social machine is concerned all we need is an understanding of the types of behaviour that occur within the machine (which may be constrained by its structure), and perhaps also information about regularities detectable within them.

Building Blocks of Social Machines

To fill out and explore this metaphor of the digitally enabled social interaction as a machine, a number of concepts must be brought into play.

Heterogeneous Human-Computer Networks

The Web, *qua* graph of links, is developed around networks of heterogeneous agents, some of which are technological, and some of which may be non-human legal entities (for example companies, charities, organisations), as well as all their human participants. ANT considers these networks as fundamental, and resists methodological distinction between humans and non-humans (referring to them all as *actants*). However, as we will see in the next section, this methodological principle suppresses the entire imperative to juxtapose the word ‘social’ with the word ‘machine’—Latour’s idea of an actant is somewhat (perhaps deliberately) paradoxical (Latour 2005) in any case, but ANT is open to an actor network containing no humans at all, whereas this would hardly count as a social machine without some explanation of how it reached that state. Such networks are, however, highly heterogeneous, containing both old and new pieces of technology (designed for different circumstances and purposes), and human agents with their varying interests, skills and resources. These networks coalesce around particular interests or topics, whether it be encyclopaedia-making, open data or education.

The networks which underlie social machines are fluid, transient and constantly remaking themselves, and the interests of the network elements, human and artificial, are in constant flux, responding to changes in the network which themselves might be driven by changes to its elements (Tinati et al. 2013; De Roure et al. 2015), as well as exogenous factors beyond the social machine. New elements will join, or attempt to join, the network if they support its agenda, and the agenda might change as the elements change. Other elements may drift in and out of the network, perhaps providing a necessary tool or application to push the agenda forward. A static representation of such a network can only be a snapshot or an approximation of a fleeting state in time, as the network is fundamentally dynamic. Agents will typically be members of several networks, online, offline and hybrid.

Meanwhile, networks will be detectable at various scales, and it is not unlikely that focusing in on an element of a large-scale network (opening up a black box) may itself expose another heterogeneous network at the smaller scale; in other words, many networks will be networks of networks, and we may find that the network structure is fractal. The social machine does not prefigure or structure these interactions—rather it is created by them, and has meaning only as coherence through time. It may also be that some coordination mechanisms can only work at particular scales, and in particular that some structures may require large scales to function; such structural determination has been observed in other cooperative contexts (e.g. ant colonies—Planqué et al. 2010).

It should also be emphasised, as we hinted earlier, that there is nothing specific to social machines as we conceptualise them that requires computer technology to be in place. The *Oxford English Dictionary* was compiled, from its beginnings in 1857, with the help of volunteers responding to appeals for examples of word usage (Winchester 2003), and of course these have only recently been digitised (http://public.oed.com/appeals/). The dictionary itself can be thought of as a social machine, and since its basic contributory structure has not changed in principle we have to agree that if it is a social machine now, so it was in 1857. As communications technology, and the practices exploiting it, have developed, the possibilities of collaboration at a distance have fired imaginations. In 1910, the arrest of Dr. Crippen and Ethel Le Neve for the murder of his wife, orchestrated via wireless telegraphy when he was recognised on board a ship bound for Canada, was a kind of policing social machine involving Scotland Yard, the Canadian authorities, the crew of the ship, the White Star shipping line which arranged for a detective to overtake Crippen, and not least the media which had publicised Crippen's and Le Neve's story and photographs. The mass media was an important development to increase scale; an interesting detective novel, *Thin Air*, by Howard Browne (1954) concerns a Madison Avenue advertising man, coincidentally also suspected of murdering his wife, who uses his publicity skills to crowdsource an investigation into her whereabouts.

However, the affordances of digital networked technology enable a quantitative improvement in scale that surely qualifies as qualitative. Computers connected by the Internet and Web not only support growth, but also promote heterogeneity (partly by increasing the pool of potential participants, and partly by making it easier to collaborate), and may also support longevity by providing a ready infrastructure for the administration of the network.

Social Computing

Over the period of the Sociam project, the phenomenon of social computing has increased significantly; the exact phrase 'social computing' produces about 9000 results in a search in Google Scholar limited to 2012–2013, up to 14,000 results in 2015–2016. Social computing itself, as the example of the DARPA balloon challenge implies, involves an interaction between digital systems and human systems. Such interactions are premised on the observation that many human interactions involve some information processing, the transfer of information or the provision of feedback. Where this information is socially produced, we have a type of social computing (Robertson and Giunchiglia 2013).

Digital technology can support this type of social computing in a number of ways, in particular by facilitating human communication, augmenting communication or using data about interactions to create new information. Technologies such as mobile computing or the cloud, as well as data about locations, e-commerce transactions, or even ease or difficulty with which a problem (e.g. of translation) is performed, are all helping embed the computing within a social matrix (Shadbolt et al. 2016). Depending on how information is aggregated, social computing may help reveal what Surowiecki (2004) has called the *wisdom of crowds*. Crowd wisdom is available when a large and diverse enough group can aggregate its opinion effectively to avoid numerous common biases, such as groupthink.

The digital technology may also, as well as facilitating communication, be the locus of interaction. In the case of the construction of some socially or economically valuable artefact—such as a Wikipedia article (see below)—the information that is passed between participants is the current state of the artefact, or alternatively the artefact itself. The appropriate wiki is sufficient to host the interaction, and to reveal the current state of the article's edit to all participants. The interaction, for example via a social networking site, could bring together a group of people that is either larger than would be feasible in pre-networked days, or too geographically remote. It may be able to improve the interaction either by introducing asynchronous communication or alternatively by facilitating synchronous communication that would not be possible otherwise. Another example of facilitating and hosting communication would be an online auction.

The information produced via facilitated communication need not be provided to the participants themselves; this type of one-way interaction includes crowdsourcing. One example here is reCAPTCHA (von Ahn et al. 2008), where information about what unclear texts may mean is provided by a crowd to Google which aggregates it and uses it for its own purposes; the members of the crowd only interact with the system as a means of proving themselves human to a website that is using reCAPTCHA, and they do not care about the problems they solve in order to do so. Or it may be, as with examples of citizen science (see below), that the information may be fed back to participants to improve their experience or incentivise their participation, although its main purpose is to facilitate the production of scientific knowledge.

Examples of social computing as augmented communication include reputation systems, reviewer ratings or numbers of 'likes', where the augmentation consists of gathering and aggregating feedback. Clearly the borderline between facilitated and augmented information is blurry.

The creation of new information could involve complex algorithms, such as collaborative filtering to produce suggestions of books that similar purchasers have rated positively, or Google's PageRank. In such cases, actual digital computing, which could not currently be conceived without the use of sophisticated computers, is embedded into the social computation. The information may be presented in a particular way to help support social interaction—for instance in a gamified environment, participants may perhaps only see motivational information such as their own ranking in terms of the value of their participation, thereby incentivising them to 'play harder'.

In this book, we will develop our understanding of these types of mechanism, which will feed into the mechanistic explanations of social machines' behaviour discussed above.

Communication

The interest of social machines is dependent upon their being social. We have already noted that the behaviour of individual participants is not necessarily explained by the social machine, or by the same explanatory techniques. A key factor in locating the social elements in a social machine is to look at patterns of communication.

Communication and Engagement Within Social Machines

The sociality of social machines can be divided into three classes, which we can set out in increasing order of interest as follows. Firstly, there are systems where the social interaction of the participants is carried out by the technology, irrespective of their intentions or even their knowledge. For example, in reCAPTCHA, the machine first decides whether a participant is human, and secondly if so uses the participant's input to help determine the interpretation of a disputed image. A non-social version of crowdsourcing, where the crowd merely solves a problem (for example a search ranking), is another example. There is clearly no social interaction here as intended by the individual participants. Indeed, in that case, relatively few of the participants even know there is a technologically mediated interaction with other humans going on. Such machines reverse Berners-Lee's original idea, as it is the humans doing the administration, and the creative work is done by the machine. One is reminded of Aristotle's discussion in the *Politics* of slaves as living tools (1.1253b).

The second class of social machine contains those machines where the behaviour is socially constituted and understood to be so by the participants, but where the interactions may not always be person to person. Crowdsourcing systems may be understood as such by participants, and the identity as a member of the social machine (e.g. a Mechanical Turker, or a participant in a Zooniverse citizen science project) may be important to the participants. They may interact on that basis, even forming organisations or trade unions to strengthen their position vis-à-vis the social machine. But any engagement that they do have with each other happens incidentally to the interactions that matter to the social machine as a whole.

The third class of social machine contains those machines in which behaviour is not only socially constituted, but also requires engagement between the participants. An example here would be Wikipedia, where the social machine consists of the incremental editing of the work of oneself and others. The work is visible to all and its creation may well require further interaction (conversations, debates, arguments and even edit wars), all within a structured social setting based on authority.

The boundaries between these classes are porous, and different (and often unanticipated) types of communication will typically be encountered. For instance, in its strict guise, Mechanical Turk pays workers to perform a cognitive task that is either impossible or hugely expensive for a computer. The worker may have no knowledge of the larger task or what they are annotating. A computer pumping a human for information is as unsocial as a human pumping a computer. Yet the former situation isn't always at equilibrium. At first blush, it has all the 'sociality' of a sweatshop in which the employees have no communication with each other and little understanding of the end product. Yet crowdworkers do have an identity, and make strong efforts to undermine the invisibility that seems to underpin the popular perception of crowdsourcing in general and Mechanical Turk in particular. Although Mechanical Turk offers no means of communication between 'Turkers', several unofficial discussion forums have grown up out of the strong desire to communicate, build trust, gain experience and expertise, and form communities, including Turker Nation, MTurk Forum, MTurkGrind and MTurkCrowd, each of which has its own distinct style (Martin et al. 2014). Activist technologies such as Turkopticon allow workers to publicise and evaluate their relationships with employers, and provide common infrastructure to enable workers to engage in mutual aid (Irani and Silberman 2013). Communication between Turkers emerges from the needs of a de facto community subtended by the original anonymous crowdsourcing platform.

As another example of this unbargained-for sociality, we can consider Facebook's ill-fated attempt to produce a participatory governance system for its privacy policies in 2009, ultimately scrapped in 2012. Under the scheme, Facebook would put up a proposed change to its privacy policies, and if 7000 comments on the policy were received, the change would be put to a vote of all Facebook users within a week. If a majority voted against the change with a turnout of 30%, then the change would not be adopted; turnout of less than 30% would result in a non-binding vote. Of course, such a turnout would require hundreds of millions of voters, and so the threshold was never reached (indeed, turnout never reached 0.5%, and was usually less than 0.1%). Each vote under the regime had a clear majority (never less than 74–26%), which in all but one case was ignored. In the end, Facebook decided the scrap the system, triggering a mandatory vote on the proposal to end democracy, which produced an 88–12% vote across 668,000 voters to preserve the system, a large number but equating to a turnout of 0.06% (https://newsroom.fb.com/news/2012/06/the-facebook-site-governance-vote/). The vote was therefore non-binding, and the experiment ceased.

Facebook presented its policy proposals as take it or leave it, not allowing users to discuss or shape them; gave very little time for support or opposition to mobilise or campaign; and required users to carry out a series of complex and mildly tedious activities (read and understand the policy changes, write a comment on a separate page and download an app to cast a vote). Campaigner Max Schrems set up a website, our-policy.org (now defunct), to coordinate responses to Facebook, and many voters argued in their comments to Facebook that they wanted to move discussion onto our-policy.org to facilitate campaigning, discussion and even the vote itself. In the end, Facebook resisted, or ignored, these comments, and the participatory element in its privacy governance died (for a detailed analysis of this forgotten but fascinating episode, see Engelmann et al. 2018).

This classification of social machines in terms of sociality and engagement connects with Haythornthwaite's (2009) binary distinction between lightweight and heavyweight systems of peer production. Lightweight crowdsourcing is based on microparticipation from unconnected individuals, while the heavyweight model assumes strong connections among a committed set of members. Dimensions such as task interdependence, authority control and group focus help determine whether the result is lightweight weak-tie attachment to a common purpose, enacted through authority-determined, rule-based contribution, or heavyweight strong-tie affiliation with community members and community purpose, enacted through internally negotiated, peer-reviewed contribution. As the discussion in this section has hinted, the position is more nuanced than binary, but Haythornthwaite pinpoints many of the important determinants of the type of social machine that result from the digitally mediated interaction.

The interest of social machines depends on the importance of the social interaction, and indeed we might hypothesise that the most successful social machines are ones within which people choose to communicate. Therefore, in this book, our major focus will be on machines which facilitate, or require, human-to-human communication. To be of interest as a social machine, its behaviour should to a significant extent be socially constituted. Social interaction must provide a significant proportion of the embodiment of the social machine's actions, and the heart of those interactions is communication. It will be difficult to draw a hard-and-fast line between social machines with interesting social interactions and those, like reCAPTCHA, without—even harder because of the constant flux of technology and practice. However, it is fair to say that the first class of social machines, where the people-to-people communication is minimal to non-existent, is probably better studied using ANT or standard network analysis techniques, rather than using the methods and tools which we will discuss in this book; little is gained by engaging with them as social machines.

Communication as Annotation

Without wishing to make grand claims, it is reasonable to hypothesise that part of the novelty of social machines stems from novel infrastructure they supply for communication. The technology of the social machine involves the creation of some kind of representation, framework or graph of relationships, and communication takes place by adding content to this representation. In other words, in many instances at the least, communication is a process of *annotation*.[2]

For instance, on a social network such as Facebook or Twitter, the underlying framework is a massive graph of social relationships (friends, following, etc.). Messages (posts, tweets, etc.) are attached to that graph, which is an essential part of how to access or participate in communication. In a citizen science machine, or Wikipedia, objects such as representations of scientific phenomena or encyclopaedia articles are annotated to share insights and to increase knowledge incrementally. Such technological mechanisms can be overlooked, but have been essential to the technical management of all sorts of scalable collaborative knowledge projects for some time. The structure to be annotated is publicly available and searchable; this facilitates the creation of diverse, distributed and geographically dispersed social machines by supporting discovery by remote actors, and ultimately therefore is a mechanism by which social machines can achieve scale.

Thinking of communication as annotation has value as it exposes a number of issues with such a mechanism, which research into annotation methods and tools has begun to address. As one example, consider the Distributed Annotation System (DAS) of the Human Genome Project (Dowell et al. 2001), which itself evolved into a community-driven project called BioDAS (http://biodas.open-bio.org/wiki/Main_Page).

[2] Thanks to Peter Buneman for this idea; we draw upon his unpublished work in this section, and also in Chapter 2.

DAS enabled scientists to annotate genetic sequences from their own visualisation tool, so that it could be seen in alternative tools, which illustrates that many people who wish to communicate across a community which is mediated by some framework or structure will be using different devices or tools, and will need to be able to not only write to the structure, but also read the contributions of others. They raise issues both technical (how should annotations propagate through edits or queries?) and social (how do you deter, detect or remove social abuse?). In particular, if we follow through the idea that communication within a social machine is a type of annotation, then we are drawn to the following four research issues:

- Annotations require some kind of data structure, framework or frame of reference which can be used to attach the annotation to some data. How do we specify this anchor point? Does the annotation depend on the particular view of the data, for example, thinking of a view in database terms as a result set produced by a specific query?
- Is there a difference between the annotations and the data being annotated? Taking the example of a database, if a column is added to a relational database in order to hold the annotations to the rows of the database, that does not affect the rest of the database. The result is a larger table in which there is no principled difference between the annotations and what they annotate. In a discussion group or forum, the basic operations are starting a topic thread, and replying to another message within a thread (i.e. annotating). The result is a tree of annotations to a root. Annotation is perhaps best seen as a relationship between data elements, relative and not absolute (Buneman et al. 2013).
- How are annotations separated from the underlying data structure? In general, annotation systems are intended to leave certain existing properties of the original data intact. A natural way to enforce this idea is to put a firewall between annotations and the data, to keep a separate database for the metadata with links to the original, to prevent applications which modify the original having access to the metadata and to prevent the annotating agents from modifying the original. This, however, loses a lot of the fluency and flexibility that should characterise social machines, and so we should look for operations and applications that act on a unified data structure, which themselves do not allow annotations to have undue influence.
- Access to a database is via a query language, and hence it is important that annotations propagate through queries where the database underpins a social machine. If we see this propagation as an influence on the outputs of queries, there is an apparent tension with the requirement that annotations don't influence the data. Annotations can be separated from the data in a query, but a problem of reattachment is then raised, and in any case that creates an apparent tension with the requirement to treat annotations and data as relational rather than absolute (Buneman et al. 2002; Green et al. 2007).

Annotation is an intriguing idea for understanding communication within social machines, and many of the structural issues involved raise questions and tensions. The more integrated and less hierarchical the social machine, the more important it