Social Machines in Practice: Solutions, Stakeholders and Scopes

Clare J. Hooper
IT Innovation Centre
University of Southampton
Gamma House, Enterprise Road
Southampton SO16 7NS, UK
cjh@it-innovation.soton.ac.uk

Brian Bailey
Uncommon in Common
PO Box 2682
Austin, TX 78768
bb@uncommon.cc

Hugh Glaser
Ethos Ltd., Seme4 Ltd. & University of Southampton
White House, Beech Road
Merstham RH1 3AE, UK
hugh.glaser@ethosvo.org

James Hendler
Centre for Creative Computing
Bath Spa University, Corsham Court
Corsham SN13 0BZ, UK
hendler@cs.rpi.edu

ABSTRACT
This paper frames social machines as problem solving entities, demonstrating how their ecosystems address multiple stakeholders' problems. It enumerates aspects relevant to the theory and real-world practice of social machines, based on qualitative observations from our experiences building them. We frame evolving issues including: changing functionality, users, data and context; geographical and temporal scope (considering data granularity and visibility); and social scope. The latter is wide-ranging, including motivation, trust, experience, security, governance, control, provenance, privacy and law. We provide suggestions about building flexibility into social machines to allow for change, and defining social machines in terms of problems and stakeholders.

Categories and Subject Descriptors
H.4.m [Information Systems Applications]: Miscellaneous

Keywords
Social machines; stakeholders; linked data.

1. INTRODUCTION
This paper focuses on social machines, a socio-technical construct by which a human-machine collective achieves greater things than possible of the individual 'parts' working alone. Berners-Lee [2] defines social machines as 'processes in which the people do the creative work and the machine does the administration'. Classic examples of social machines see people collaborating to produce content (e.g. Wikipedia and Galaxy Zoo), but examples include tools and counter-tools (e.g. spam networks and reCAPTCHA [1]) and subsets of social networks such as Twitter bots [12].

This paper provides two contributions to the state of the art:

1. A lens for viewing social machines as problem solving machines, showing how their ecosystems converge to address issues relevant to multiple stakeholders;
2. Enumeration of what aspects to consider – and why – when building a social machine, providing a research contribution that is relevant to real-world practice.

Section 2 introduces social machines and stakeholder analysis, and Section 3 frames social machines in terms of problem solving and stakeholders. Section 4 discusses geographical, temporal and social scope, and change. Section 5 looks at building machines.

2. BACKGROUND AND RELATED WORK
2.1 Social Machines
The concept of social machines for the Web was proposed in 1999 [2]. Early work reflects on social machine interactions, lifecycles and the Internet of Things [12], asking whether social machines may be distinguished by purpose: we build on this topic in Section 3. Other work, relevant but parallel to this paper, discusses social machine ecosystems, focusing on observation methods [13].

[5] discusses Semantic Web challenges for social machines, noting the need for fine-grained information access and control, and context mechanisms. Physical, temporal and social scope are a key part of the latter, and are discussed in Section 4.

Hooper et al describe the process of building a social machine, an online TimeBank (a time-based way for people to give and receive services) [7]. They found implications for social machine design regarding the primacy of experience and values, and raised the importance of presentation and granularity of spatial and time data. They identified key questions for building a social machine:

- How to ensure experiential aspects such as fairness and reciprocity, trust and security online?
- How to foster connectedness and community?
- How may the rules of a social machine need modifying over time, as that machine grows in size or changes in cultural composition? (Hendler and Berners-Lee refer to this as *governance* of the social community.)

WebSci '16, May 22 - 25, 2016, Hannover, Germany.

ACM 978-1-4503-4208-7/16/05…$15.00.
DOI: http://dx.doi.org/10.1145/2908131.2908176

Such topics have also arisen in the SOCIAM project, with a talk on crowdsourcing and open data[1] noting that successful volunteer crowdsourcing (a specific type of social machine) is hard to predict or replicate, much like the success of social machines in general. As with the TimeBanking project [7], it was observed that building a community from scratch is not trivial, and that crowdsourcing reward models are often easier to study and control if performance can be reliably measured.

2.2 Stakeholder Analysis

Most if not all issues that social machines respond to have multiple stakeholders. There is a wealth of literature on stakeholder analysis; we focus on the context of social machines.

The SEQUOIA method [11] is a tool for socio-economic impact assessment of projects that uses stakeholder analysis. It involves analysing the area of impact, the baseline scenario (i.e. the scenario without intervention), the scenario with intervention, followed by quantification and impact assessment. Stakeholder identification is essential for an accurate measurement.

SESERV's Tussle Method [9], from the context of future internet technologies, is very relevant: its goal is to understand issues such as control, contention, repurposing and responsibility among multiple stakeholders. It aims to maintain balance between stakeholders without prejudice to other interested parties, and involves identification of stakeholders and 'tussles' (areas of contention) between them. For each tussle, the practitioner analyses the distribution of control, assesses impact where this is out of balance, and identifies how to rebalance it if necessary.

3. PROBLEM SOLVING ECOSYSTEMS

3.1 Problem Solving Social Machines

As described in Section 1, examples of social machine include Wikipedia and Galaxy Zoo (where people collaborate to produce context), spam networks and reCAPTCHA [1] (tools and counter-tools), and Twitter bots (subsets of social networks). Open source software communities are rarely considered in this context, but can be viewed as highly effective social machines.

We argue that a social machine is more than just an online community; social networks such as Twitter and code hosting and community services such as GitHub are really platforms that enable social machines. Twitter supports networks of bots [12] that work towards particular goals, whether related to spam, spam reporting, humour, fact checking, or some other topic. Indeed, a useful framing of a social machine is to ask: "what problem does this solve?" The problem may be physical, digital or hybrid: for example, a TimeBank may support both physical and digital exchange [7], but other social machines are purely digital (e.g. spam and counter-spam bots on Twitter). In all cases, successful social machines amplify power. Table 1 describes a set of diverse social machines alongside some of the problems that they solve.

The multifaceted nature of social machines, and their amplification of power, means they can be a highly appropriate response to 'wicked' design problems: such problems exhibit great complexity, often involving changing, incomplete or conflicting requirements, and frequently being entangled with other big issues [3]. Such problems vary in scope, with examples including those mentioned in [5]: global problems such as climate change and curing cancer, and local problems, such as creating a community to address a social issue in a city suburb.

Table 1. Some social machines and problems they solve

Social Machine	Examples of Problems Solved
eBay	Allow consumers to buy and sell goods to a very wide audience of fellow consumers.
Reddit	Filtering news and discussion topics, and rapid community generation via subreddits.
Gowalla	Cataloging significant and mundane locations throughout the world with geolocation data, text, photos and artwork.
Facebook	Incidental and significant knowledge about the activities of family, friends and acquaintances; a digital experience not unlike that of chancing upon one's contacts when living in a village. Of special note is how communities occasionally use Facebook for coordination, such as in the "Arab Spring" protests in Egypt.
SkillsPlanner	An open data approach to link people's skills in a workplace market, connecting supply and demand.

Social machines vary widely in intentionality. Some machines are created by a single stakeholder who aims to solve a specific problem: Galaxy Zoo, created by scientists for people to help solve the problem of galaxy classification, is an excellent example. Other machines, however, are more unplanned and emergent, typically run by communities: one example is human flesh search, which originally emerged from Chinese chat groups and is typically but not always used to solve a real-world problem, using a crowd-sourcing type of solution. Online communities can thus be a source of emergent social machines, with for example birder and mushroom communities producing such artefacts.

3.2 Social Machines and Stakeholders

The person, community, or organisation to initiate or support a social machine will have an objective in doing so, but it is highly likely that this is different from the objective of the end users of the machine (spam networks are an obvious example of conflicting goals by different stakeholders). The emergent nature of social machines and their liability to be used in unexpected ways compounds this important aspect. From this, we can see that although any social machine will have at least two rulesets (one for the computers and one for the humans), given the existence of multiple stakeholders, a machine is likely to have multiple rulesets. Table 2 describes problems solved by a small selection of previously-described social machines for some of their stakeholders.

As illustrated by SkillsPlanner, a range of stakeholders exist, some of whom are directly involved in the social machine and many of whom are not. Other stakeholders not included in the SkillsPlanner example include brokerages, local people, politicians and professional bodies. A social machine to address all of these issues would resolve some very complex tussles, representing often-conflicting interests from a diversity of sectors. The owning company (Ethos)'s collaborative approach to this issue is to explicitly contact stakeholders and get their buy-in to being involved, then combine all relevant data to approach the issue. As experience in collaborative work practices and the social machines that enable them becomes wider, we expect the social machine design to improve and become better at satisfying the multiple stakeholder needs.

[1] Simperl, "Open Data and Social Machines" at IT as a Utility Community Conference, Southampton, UK, 7 July 2015.

Table 2. How some social machines address stakeholder aims

Social machine	Stakeholder	Stakeholder's Objective
Facebook	End user	Incidental and significant knowledge about the activities of family, friends and acquaintances (other aims exist).
	Shareholder	Make money.
	Advertiser	A platform to deliver ads to people.
	App developer	A platform to deliver apps to people.
Skills-Planner	Ethos (owning company)	Make money and create social good (address London's construction skills shortage, ensuring there is a workforce with appropriate skills to build in London over coming years).
	Further Education colleges	Employment for their graduates.
	Local Council	Employment for inhabitants; long-term value from building work (rather than migrant labour that goes elsewhere).
	Employees	Work.
	Projects	Healthy construction companies they can contract
	Construction companies	Sub-contractors that can deliver.
Gowalla	End user	Record travel via virtual passport, discover new places, follow friends.
	Shareholder	Make money.
	Local business	A platform to promote specific business locations.
	Advertiser	A platform to deliver ads to people.

A next step from this point is to identify the relevant scopes for the social machine, and its lifecycle. Ideally, of course, designers of social machines would be able to predict their evolution, but emergence is inherent to social machines and makes such prediction extremely difficult if not impossible.

Even a single, apparently simple social machine is likely to solve multiple problems for very different stakeholders. The following section considers aspects in the design of a social machine in addition to the problem being solved and the stakeholders.

4. SOCIAL MACHINES IN CONTEXT

The SOCIAM project suggests four categories of social machine[2]:

1. social networks (e.g. Twitter and Facebook),
2. knowledge sharing (e.g. Wikipedia and Reddit),
3. public service (e.g. eBay and Fix my street),
4. citizen science (e.g. Galaxy Zoo and Shroomtalk).

This diversity of interacting social machines makes it difficult to delineate where one machine ends and the next begins. Doing so is important in designing and engineering machines: of course, not all successful social machines were planned, but a strong theoretical basis can support their design and implementation.

[2] See http://sociam.org/social-machines for more examples within each category.

This section explores key aspects, in addition to problems solved and stakeholders, for understanding and building social machines: geographical scope, temporal scope, social scope and change.

4.1 Geographical Scope

For some social machines, geographic location is central: consider social machines to catalogue locations (e.g. Gowalla, Foursquare), enable consumer-to-consumer reuse of goods (e.g. Freecycle, eBay), or which involve meet-ups (e.g. TimeBanking, Tinder). In some cases, location can be of variable importance: for example, eBay auctions can be set up for international postage, national postage, or local collection only.

Location is still relevant for social machines for which end user location is not central: people are more likely to join social networks that host their friends, who are likely to be from a similar area of the world. Facebook's early model of allowing users to sign up on a campus-by-campus basis reflects this. In addition, timezones are relevant for interactions between users, whether in terms of allowing a rapid asynchronous back-and-forth as in interactions on Twitter, or of allowing real-time interactions as in Uncommon's "Table for Six" phone calls.

Questions on how users share geographical data include:

- How much detail (e.g. country, city, home address)
- How public (e.g. public, visible to some users, private)
- How frequently (e.g. one-time update when setting up a profile, regular updates when checking into locations, continuous broadcast via GPS)

Geographical aspects are often complicated to implement: for example, if offering a UK user of a system the option to specify a location, it could be in terms of country, county, town, postcode, road, parish or any of a number of other options. In India, where addresses are not standardised, the problem is still harder: for example, a valid postal address can include driving directions from a known location. Careful thought is therefore required when building the interface of a geographically-aware social machine.

Wider issues also apply. As has been noted in the context of TimeBanking, people's travel abilities and requirements are unlikely to fit neatly into a list of districts: someone with certain physical needs may only be able to use particular trains and be unable to manage unassisted access to some towns altogether [7].

Aspects such as continuous broadcast via GPS, typically seen in fitness apps that track running, bring the Internet of Things and other cyberphysical systems into the picture: social machines are by no means limited to the web.

Of course, concerns relating to privacy and data security come into their own in this topic: this is discussed further in Section 4.3.

4.2 Temporal Scope

As with physical scope, temporal scope is relevant to almost every social machine, but is much more central to some. Social machines for which time is a crucial aspect include crime, traffic and news reporting, as well as machines that facilitate exchanges, e.g. both eBay auctions and TimeBank offers will expire. Issues of timeliness still apply elsewhere, for example in terms of providing vital context for online data as well as provenance.

As with geographical information, temporal data's granularity is relevant. Many social machines timestamp content to the nearest minute (e.g. news stories or social network posts), or use no timestamp (e.g. citizen science machines such as Galaxy Zoo, or public service machines such as Fix My Street). Other social machines use different granularities: for example, sensors for weather or parking data may report every five minutes, while also making hourly, daily and weekly predictions. As with location, the interface for time reporting can be difficult. For example, a birdwatchers' social machine (birdforum.com) reports the time a blog is posted automatically, but blog posts often report sightings from different times, and in non-standard reporting formats ("A few weeks ago while traveling through Belgium…").

The interaction between spatial and temporal data is of note, particularly in the context of privacy: people are often happy for spatial data's accuracy to increase as temporal data's accuracy decreases (Alice may be happy to publish an address from 20 years ago, but not a current address, or her GPS co-ordinates on a run earlier today, but not while she is doing it).

4.3 Social Scope

A huge number of social factors are relevant to the success or failure of any social machine. There is an extent to which social scope is derived from geographical scope, but rather than physical location it is concerned with the following aspects, relating to:

- Social expectations from humans: motivation, homophily, trust, experience
- Technology: data security, governance and control
- Social machines as a whole: provenance, privacy, international law

The above bullets assume boundaries that do not always exist, i.e. data security and governance (technology) relate to trust and experience (humans) and privacy (social machines as a whole).

4.3.1 Social scope and human aspects

Social machines depend on contributions from motivated human participants. Motivation and incentives in social machines has been the subject of study, with [14] identifying sources of intrinsic motivation (fun; accomplishment; gaining or sharing knowledge; being social; philanthropy towards a specific group; philanthropy towards society). Thus far, extrinsic motivation for engagement with social machines has only been studied in the context of citizen science [15]. There is much debate over what motivates users of the Chinese Human Flesh Search[3]. Such work is beyond the scope of this paper, but we note that cultural similarities are likely to be highly significant to people's motivation to contribute.

Homophily is the concept that people with like traits will tend to cluster together [4]: shared languages, interests, and culture bring people together, and many social networks demonstrate this.

Trust is essential to the uptake of social machines: [7] suggests ways to bolster trust including snowball recruitment (using existing FOAF networks), regional launches (acquainting users with one another), and verification of user status with digital badges on profiles. For more on trust, in general and in the context of online technology, see [8]. (Trust also relates to the thorny issue of online identity, which is beyond the scope of this paper.)

It has been argued that experiences should be treated as first class objects in social machines [7]. User experience includes but is not limited to whether the interface is usable, encompassing social and emotional aspects [6]. For example, the discovery in Galaxy Zoo of a previously unknown astronomical phenomenon was facilitated by a user's ability to report her experience on the site.[4]

4.3.2 Social scope and technological aspects

Data security links with trust, privacy and international law: numerous data breaches have made headlines of late, leading to falls in user trust. Linked with this topic is how to convey to users when data is held securely (and, conversely, when it is not).

Good governance is essential to ensuring that a social machine functions as it should, and is thus linked with control of system use. Both are difficult topics: social machines' emergent nature means governance mechanisms may not always be included due to unpredictable turns their use (consider the rise of spam email), while closely controlled systems are unlikely to support sustainable growth. Further, as communities grow, governance issues often change, implying the design of social machines must allow for growth and change over time.

4.3.3 Social scope and social machine aspects

Social machines can show information's provenance, privacy and flows, but this rarely happens (e.g. social network users can rarely see who has viewed their posts). Transparency can increase trust, helping users make better-informed decisions about sharing [10].

Social machines must obey international law, treating legal rules as first class objects [5]. This is increasingly challenging: laws that govern aspects such as privacy, data reuse, etc. change over time and across regions – again arguing for the need for flexibility in social machine design, as we discuss in the next section.

4.4 How Social Machines Change

It is very difficult to predict how a social machine will evolve over time. As has been previously discussed, social machines have wildly differing lifespans, from years (e.g. LiveJournal), to days (#UKsnow) to hours or minutes (Twitter spambots) [12]. Social machines can change in different ways:

1. Changes to functionality, underlying technology, interface, or governance
2. Growth or decline of number of end users
3. Change in composition of end users
4. Accumulation of historical data
5. Changes in context of social machine
6. Co-evolution in concert with other social machines

The proprietors of a social machine may choose to make explicit changes, perhaps to functionality (as when LiveJournal added the ability to repost blog posts to other social networks), underlying technology (as when Facebook adopted a just in time compiler for PHP), interface (as when Twitter added a search sidebar to its homepage, making access to an existing function easier) or governance (such as when Wikipedia changed to include multiple levels of editors). These changes are controlled by the social machine's 'owners', but their impact may not always be obvious.

The number of end users of a social machine is one of the easier changes to track, and for many social machines, user growth is the key goal. In social networks, a large user group leads to increased revenue, while citizen science social machines can achieve faster results. Social machines may aim at different rates of change: Uncommon includes design decisions intended to limit addictiveness and virality. The site is closed one day each week, user visits are limited, and access is restricted to members. Rather than pursue rapid growth to attract advertisers and suitors, Uncommon's aim is a small, high-quality, sustainable community.

It is more difficult to track change to composition of end users, as when a social machine expands to encompass a greater user group from more diverse backgrounds. Facebook's roots were in

[3] https://gigaom.com/2011/08/14/are-we-the-web/

[4] https://en.wikipedia.org/wiki/Hanny's_Voorwerp

American college campuses, but as it grew and opened its doors to sign-up from anywhere, the background and average age of its users shifted. Opening a social machine to public sign-up is a significant relinquishment of control [7], although mechanisms exist to keep some control. One example is giving existing users "tickets" for sign-up to share with their friends.

As an active social machine ages, its dataset increases. This is more relevant to some machines than others: eBay's historical data is of some interest to economics and scholars, but the majority of its users are far more interested in current auctions (and very recently past ones) than anything else. The past data of citizen science social machines is, of course, of continuing interest to the owners of those machines, although less so to active users: this returns to our earlier reflections on different stakeholders.

No social machine exists in isolation, and the broader context around any machine will be changing in ways that may also affect it. The increase in people using smart phones with GPS technology greatly increased the number of possible end users for location-based systems such as Gowalla, while changes in economic circumstances leave consumers more or less likely to use services such as Uber, which allow access to vehicles for hire.

Relatedly, social machines exist in concert with one another as well as their broader surroundings. They may interact in a number of ways, in some cases begetting more social machines, in other cases competing, and in other cases cooperating. The Twitter platform is a backdrop for an example of social machines begetting one another: i.e. the social machine of Twitter spambots led to the social machine of bot reporting and blacklisting. Competition can be seen in social networks that compete for the membership and activity of a finite number of potential users. We expect the current Smart Cities initiatives to provide examples of many co-operating social machines, as the emerging social machines in local contexts, such as crime fighting, environmental maintenance, retail and transport gain power by interacting with each other, and even form new social machines.

5. BUILDING SOCIAL MACHINES

This paper has explored the diversity of social machines, describing how they respond to problems and cater to the (sometimes conflicting) needs of multiple stakeholders. Key issues include the granularity, visibility and frequency of geographical and temporal data (and the relationship between these two types of data), as well as many social issues including trust, experience, security, governance, control and law.

We note the paramount importance, when designing a social machine, to allow for growth and change over time. Internal communities will evolve alongside external context, such as related social machines and international law. Owners of social machines can control some limited aspects of change, but by no means all: we advise them to capture relevant data (about both machines and their context), and to monitor the impact of planned and unplanned changes. Users can often provide valuable insight, particularly when machines evolve in unexpected ways.

We have observed that it can be difficult to delineate where one machine ends and the next begins. Understanding the communities served and the problems solved by a social machine offers a route to defining that machine in a helpful way.

6. ACKNOWLEDGMENTS

We acknowledge colleagues at Seme4 and Ethos, who have contributed by actively trying to realise social machines. This work is supported under SOCIAM: The Theory and Practice of Social Machines. The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/2 and comprises the Universities of Oxford, Southampton, and Edinburgh.

7. REFERENCES

[1] von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M. 2008. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science. 321, 5895, 465–1468. DOI= http://dx.doi.org/10.1126/science.1160379

[2] Berners-Lee, T. and Fischetti, M. 1999. Weaving the web: The original design and ultimate destiny of the World Wide Web by its inventor. Harper, San Francisco

[3] Churchman, C.W., Guest Editorial, Management Science, Vol. 14, No. 4 (December 1967)

[4] Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

[5] Hendler, J., & Berners-Lee, T. (2010). From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*, *174*(2), 156-161.

[6] Hooper, C. J. and Millard, D. E. (2010) Teasing Apart and Piecing Together: Towards Understanding Web-based Interactions. In: Web Science 2010, Raleigh, USA.

[7] Hooper, C.J., Nind, M., Parsons, S., Power, A., Collis, A. (2015) Building a Social Machine: Co-designing a TimeBank for Inclusive Research. In: Proceedings of ACM Web Science 2015, Oxford, UK.

[8] Hooper, C.J., Pickering, J.B., Prichard, J., Ashleigh, M. (2015) TRIFoRM Final Report: TRust in IT: Factors, metRics, Models. 2015. IT as a Utility Network.

[9] Kalogiros, C., Courcoubetis, C., Stamoulis, G. D., Boniface, M., Meyer, E. T., ... & Stiller, B. (2011). *An approach to investigating socio-economic tussles arising from building the future internet* (pp. 145-159). Springer Berlin Heidelberg.

[10] Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

[11] Passani, A., Monacciani, F., Van Der Graaf, S., Spagnoli, F., Bellini, F., Debicki, M., & Dini, P. (2014). SEQUOIA: A methodology for the socio-economic impact assessment of Software-as-a-Service and Internet of Services research projects. *Research Evaluation*, *23*(2), 133-149.

[12] De Roure, D., Hooper, C.J., Meredith-Lobay, M., Page, K., Tarte, S., Cruickshank, D. and De Roure, C. 2013. Observing Social Machines part 1: what to observe?. In 22nd international conference on World Wide Web companion (WWW '13 Companion). Geneva, Switzerland, 901-904

[13] De Roure, D., Hooper, C.J., Page, K., Tarte, S., Willcox, P. (2015) Observing Social Machines Part 2: How to Observe? In: Proceedings of ACM Web Science 2015, Oxford, UK.

[14] Shadbolt, N. R., Smith, D. A., Simperl, E., Van Kleek, M., Yang, Y., & Hall, W. (2013, May). Towards a classification framework for social machines. In *Proceedings of the 22nd international conference on World Wide Web companion* (pp. 905-912).

[15] Tinati, R., Luczak-Roesch, M., Simperl, E., & Shadbolt, N. (2014, June). Motivations of citizen scientists: A quantitative investigation of forum participation. In *Proceedings of the 2014 ACM conference on Web science* (pp. 295-296). ACM.