Observing Social Machines Part 1: What to Observe?

David De Roure
Oxford e-Research Centre
University of Oxford
7 Keble Road
Oxford OX1 3QG
david.deroure@oerc.ox.ac.uk

Clare Hooper
IT Innovation Centre
University of Southampton
Gamma House, Enterprise Road
Southampton SO16 7NS
clare@ecs.soton.ac.uk

Megan Meredith-Lobay
Oxford e-Research Centre
University of Oxford
7 Keble Road
Oxford OX1 3QG
megan.meredith-lobay@oerc.ox.ac.uk

Kevin Page
Oxford e-Research Centre
University of Oxford
7 Keble Road
Oxford OX1 3QG
kevin.page@oerc.ox.ac.uk

Ségolène Tarte
Oxford e-Research Centre
University of Oxford
7 Keble Road
Oxford OX1 3QG
segolene.tarte@oerc.ox.ac.uk

Don Cruickshank
ECS
University of
Southampton
SO17 1BJ
dgc@ecs.soton.ac.uk

Catherine De Roure
Dept of Computer Science
University of Bath
Claverton Down
Bath BA2 7AY
cadr20@bath.ac.uk

ABSTRACT
The ambition of our Social Machines Observatory is to study the Social Machines ecosystem. As a preliminary scoping exercise we consider the observation of Social Machines "in the wild" as illustrated through two example scenarios. More than identifying and classifying individual machines, we argue that we need to study interactions between machines and observe them throughout their lifecycle. We suggest that *purpose* may be a key notion to help identify individual Social Machines in composed systems. This exercise provides a basis for later work on *how* we instrument and observe the ecosystem.

Author Keywords
Social Machines; Web Observatories; Internet of Things; Mixed Methods

ACM Classification Keywords
H.4.m. Information systems applications: Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Very many examples of Social Machines [1] are proffered by the emerging research community, from Wikipedia to Ushahidi and reCAPTCHA to Galaxy Zoo. This fuels an active debate on the definition and typology of Social Machines, and the realisation that to study them will require the choice of usefully representative examples in a rich and emerging multidimensional design space.

In order to identify what type of methods to adopt to observe Social Machines, we first need to establish what we intend to observe. In this short paper we argue for observation of multiple interacting machines throughout their lifecycles, and further suggest that the diversity and polymorphous aspects of Social Machines require the use of mixed observation methodologies in order to capture their driving forces.

We first present two scenarios which have been the basis of our discussions, developed directly from the experience of the authors. We then present some of the defining variables of their constituent Social Machines in order to better grasp what they are, before we conclude with some methodological recommendations for the observation of Social Machines.

Scenarios

1) *The Machines of Spam*
Administrators of social websites are familiar with the problem of spam accounts, whereby accounts are created for purposes other than legitimate use. To limit abuse, mechanisms like reCAPTCHA are used to ensure that accounts are created by humans.

Detailed investigation of one such attack during 2012 revealed its source: "spam as a service" websites that pay people to create accounts in order to promote products and sites. Our response to this particular attack was firstly to create a team of administrators and scripts to assist with the "despamming", and secondly to make use of a blacklisting website which keeps the details of spammers so that they can be identified during the process of account creation.

Consider this now from a Social machines perspective. The original social website is a Social Machine protected by a Social Machine (reCAPTCHA). The spam attack came from another part of the ecosystem, but again from Social Machines. To combat the problem we created a *new* Social Machine, then we evolved our original machine to made use of another (blacklisting). This example illustrates aspects of multiple *interacting* Social Machines as well as their design and composition.

2) The Befriending of RaspTreePi

In December 2012 a Raspberry Pi that had been set up to control Christmas Tree lights was connected to the Twitter API by a simple bot, enabling people to tweet commands at the tree and change the behaviour of the sequence of lights. The next day we found that our tree had a new follower. The follower was simply tweeting quotations and it seemed likely that it was a bot too. Looking at its profile, it carried a link to a dating site. The second bot soon disappeared from Twitter and we may presume this was the result of human complaint and/or automated rules triggered by frequent formation of relationships and posting duplicate or irrelevant content.

Whether or not the tree and bot is a Social Machine, or indeed the dating site, are points of consideration. This scenario illustrates a degree of automation involving machine-to-machine communication without human mediation, automatic assembly, and the presumed bot detection algorithm illustrates a built-in observation mechanism. It also raises questions relating to identity: while Social Machines do not claim that people and machines are interchangeable, here we see machines impersonating people. Indeed, people can pay for human twitter followers (perhaps buying them on the eBay Social Machine) but there is a risk that these are bots.

SOME CHARACTERITICS OF SOCIAL MACHINES

1) INTERACTING SOCIAL MACHINES

The power of the Social Machines metaphor comes from the notion that a machine is not just a computer which has some users, but rather it is a purposefully designed sociotechnical system comprising machines and people. We can then view the ecosystem as a set of interacting Social Machines, rather than a layer of computers used by a layer of humans.

Hence the study of individual machines is only part of the story: we must also study their interactions. Clearly these are sometimes mediated by humans (e.g. spam) and sometimes computers where automation occurs. Our scenarios suggest they may also be mediated by yet more Social Machines. Machines and users might simultaneously be members of multiple/different social machines at the same time, and our suggestion is that each social machine is delimited by *purpose*.

As well as explicit communication, the machines are coupled through the ecosystem in which they co-evolve: if machines are designed to attract users in order to flourish (a motivation in both scenarios) then effectively they are competing for popularity. There are limits to numbers of human users and to the attention of each, and the demand has led to provision of bots instead as fake humans.

2) LIFECYCLES & LIFESPANS

While we can study machines that function and interact today, the time axis is also crucial: how did machines come about, how did they (co-)evolve? It is likely that a Social Machine as structured at any instant will not remain so, rather it will be modified or become redundant. This is why we should think about the *design* and *construction* of machines, as well as their operation. This is contextualised by their environment. Indeed, Social Machines may exert strong influences upon one another, building whole ecosystems of which we may be unaware.

It is worth noting that the lifecycle of a Social Machine can vary hugely: from the 12 long years of Wikipedia's life to date, to the 12 days of relevance and attention that the #UKsnow hashtag enjoys, to 12 hours (or minutes) that a Twitter spam bot may last before being removed due to the anti-spam Social Machine. Relatedly it should be noted that, although some Social Machines straddle vast quantities of data and may be prominent due to their popularity, there is also a spectrum of sizes (e.g. of social network) which should be reflected in their study.

3) MORE THAN THE WEB

Common examples of Social Machines are social websites and crowdsourcing sites, especially where some aspect of the behavior is socially constructed. The focus is very much on the Web as the interface between human and machines. While many Social Machines involve intersections between the digital and physical worlds – ranging from "citizen sensing" to interpretation of archaeological digs – the second scenario hints at a more intricate coupling in emerging cyberphysical systems. We anticipate increasing prevalence of such systems.

For example, increasing attention is being given to sustainability and energy consumption[1], with some work arguably building office-based Social Machines[2]. Given this precedent along with the augmentation of houses to tweet data including (but not limited to) energy usage, it is no stretch to imagine small communities (e.g. a street of several dozen houses) that use an imprecise indicator (traffic light style colours, or grades from 1 - 5) to show who among neighbours are using more or less energy. This is an instant, small-scale Social Machine, the effects of which could scale to have a huge impact.[3]

---

[1] http://www.bbc.co.uk/news/technology-20173641

[2] http://mqtt.org/projects/andy_house

[3] Further along the "cyberphysical" axis, a Social Machine might occur within a house, wherein multiple appliances (the washing machine, cooker, dishwasher and phone charger) negotiate among themselves about scheduling access to the power grid, drawing on current grid usage data: this Social Machine, unlike the prior examples, is presided over by computers rather than humans.

Larger scale Social Machines that draw on Internet of Things rather than the Web might include systems that use car GPS information to calculate (and disseminate) current traffic congestion, or that manage shopping patterns by allowing shoppers to access live aggregate data from RFID tags as goods are bought.

AT THE CORE OF SOCIAL MACHINES

One of the fascinating aspects of a Social Machine as a (eco-)system is its ability to resist attempts at defining it. Inspect it too closely and you'll end-up studying its constituting parts in great detail (humans, machines, bots); step away too much and you loose sight of what the constituting parts are doing.

One of the main reasons why this is happening is that a successful Social Machine, whether intentionally designed or serendipitously emerging, has all the properties of an *emergent* system; the fabric itself of its constituting parts mutates under their mutual influence, as do the interactions between them. The result is that the constituting parts cannot be classified into straightforward layers anymore but actually span various layers. It follows that to observe the whole system, some methodological adaptations need to occur.

Discussions around the scenarios suggest that *purpose* is a key notion. Is Twitter one Social Machine or many socially-constructed machines with a common infrastructure? In the latter case the individual machines may be distinguished by purpose, with the caveats that (a) machines may be multipurpose (e.g. reCAPTCHA clearly has a dual role), and (b) purpose may evolve.

That individual Social Machines are constructed with a purpose is consistent with our emphasis on their design and construction. But as we know too that behaviours can be emergent, we will also be designing combinations of Social Machines in an attempt to "engineer" the desired combined and emergent behaviours within the ecosystem.

METHODOLOGICAL IMPLICATIONS

Can sufficient observations be achieved by analysing log files? The scenarios in this short paper illustrate the value of other methodologies, understanding purpose and emergent behaviours. These understandings may come about through considering the "user experience" of/in the machine, through engaging directly with the machine, and through working directly with those who design machines (successful or not).

It is essential to account for User Experience when considering these contexts. Key aspects include not only the experience of end users interacting with the Social Machine (web-based or otherwise), but also user perceptions of the system: how many people realise what Recaptchas do? How much do casual browsers of Reddit understand the importance of (or the rules governing) 'karma'? We need to understand the rules that ground Social Machines, whether elements within were designed or participate intentionally or otherwise: how does eBay's governance of sellers impact not only the experience of sellers themselves, but also buyers and other online auction sites? In the context of Social Machines, User Experience concerns not only the interaction between Human A and Computer B, but understandings and interactions between humans, computers, and social constructs / rules.

Just as the use of mixed methods is essential in Web Science as a whole, so it is the case when observing, analysing and evaluating Social Machines. If we are to answer such questions as "How do people experience the social web?", "How do we evaluate the process by which this Social Machines was constructed?", and even "How do we build a Social Machine to achieve X?", we need to be able to observe, understand and model experiences. Mixed methods help us deal with the variability of human experience by letting us access the richer insights and multiple types of finding that qualitative research methods unfold, while dealing with big quantitative data in a meaningful and scalable way. We need to use qualitative and quantitative methods *in conjunction* to triangulate and better understand our results: statistical analysis and qualitative coding yield much stronger results when the products of one inform those of the other.

MOVING FORWARD

In depth analysis of a small number of high profile Social Machines is a useful function of a Social Machines Observatory, indeed it may be necessary – but we have argued that it is not sufficient. Social Machines need to be studied within their ecosystem, especially as the purpose of that study is to be able to design and construct successful machines and this is a function of their context.

We argue then for an "ecological" perspective. The case studies in a Social Machines observatory need to be based on regions of the ecosystem involving Social Machines that are:

1. Interacting and competing with others;
2. Being designed, born and co-evolving;
3. Variable in size, purpose and intent;
4. Reflecting the trends towards cyber-physical and machine-to-machine systems.

Having chosen exemplar areas of the ecosystem we need to identify:

- The constituent Social Machines, by considering the intent and evolving purpose of constructions deemed to be a Social Machine;
- Technologies, humans and their interfaces, including the intersection with the physical world;
- Ground rules that lead to the emergent behaviour: these may be explicitly stated as rules by which people abide, encoded in how the technology works, potentially even just part of community conduct or grounded in how other Social Machines behave.

In this paper we have discussed what to observe. The next part of our work is to determine *how* to make those observations, and this will be informed also by parallel research on the definition and typology of Social Machines.

ACKNOWLEDGMENTS

We are grateful to all our colleagues for useful discussions and insights. The work of several of the authors is supported by *SOCIAM: The Theory and Practice of Social Machines*, funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1 and comprising the Universities of Southampton, Oxford and Edinburgh, and by *Digital Social Research* funded by the UK Economic and Social Research Council under grant number RES-149-34-0001-A. The "Machines of Spam" scenario arose in the development of the myExperiment social website supported by multiple projects including EPSRC grants EP/G026238/1 and EP/F05811X/1 and STREP FP7-ICT-2007-6 270192.

REFERENCES

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

[2] M. Motoyama, D. McCoy, K. Levchenko, G. M. Voelker, and S. Savage. Dirty Jobs: The Role of Freelance Labor in Web Service Abuse. In Proceedings of the USENIX Security Symposium, San Francisco, CA, August 2011.