Constructed Identity and Social Machines: A Case Study in Creative Media Production

Amy Guy
School of Informatics
University of Edinburgh
amy.guy@ed.ac.uk

Ewan Klein
School of Informatics
University of Edinburgh
ewan@inf.ed.ac.uk

ABSTRACT

Current discussions of social machines rightly emphasise a human's role as a crucial *part* of a system rather than a *user* of a system. The human 'parts' are typically considered in terms of their aggregate outcomes and collective behaviours, but human participants are rarely all equal, even within a small system. We argue that due to the complex nature of online identity, understanding participants in a more granular way is crucial for social machine observation and design. We present the results of a study of the personas portrayed by participants in a social machine that produces creative media content, and discover that inconsistent or misleading representations of individuals do not necessarily undermine the system in which they are participating. We describe a preliminary framework for making sense of human participants in social machines, and the ongoing work that develops this further.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems

Keywords

social machines; user-generated content; identity; creative media production; online communities

1. INTRODUCTION

Social machines are systems for which human and computational aspects are equally critical. They exist already as a product of the social web, on many different scales and in many different domains, often evolving, responding to technological and social developments, and interacting with each other [8, 3, 5, 15]. Human participants in social machines are often discussed in collective terms, and have yet to be given thorough consideration as individuals. They are very diverse and participate in different ways, with varying goals, motivations and outcomes. They even replicate their online presence so that they may behave in different ways according to different contexts, and may work together to construct a single image controlled by multiple people [4]. As well as conflating issues of incentives and roles, this has an impact on attribution for work done, and accountability and trustworthiness of participants in a system and the data they produce. Ignoring this can cause problems with observing social machines (eg. assuming that participants who adopt multiple identities do so with an intention to deceive) and with designing them (eg. imposing restrictions that conflict with some adapted use of the system).

In this paper we argue that due to the complex nature of online identity, understanding nuanced identity behaviours of participants in a more granular way is crucial for social machine observation and design. We briefly describe *creative media production social machines*, and present the results of a study of personas portrayed by participants in one of these. We then describe a preliminary framework for making sense of human participants in social machines, and propose extensions to the study to further develop the framework.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2578836.

2. CREATIVE MEDIA PRODUCTION SOCIAL MACHINES

Amongst the plethora of user-generated content on the web are a huge number of works of creative media, and behind these are independent content creators pushing their work to a global audience and actively seeking to extend their reach. Within this ecosystem we can observe a variety of different sizes and forms of *creative media production social machines*. The definition of creative media production social machines encompasses a broad class of systems where:

- humans may use a purely digital, or combination of digital and analogue methods, and a degree of creative effort, to produce media content;
- the content is published and publicly accessible on the web;
- a global audience may consume, curate and commentate on this content in technologically-mediated environments.

These social machines exist both within and across content host platforms (eg. YouTube) and within and across online communities and social networks. Many, if not all, media types and genres are represented among the media artefacts that emerge from these systems, and such artefacts can have a sometimes profound effect on media and culture in the offline world.

Figure 1 shows the interconnected social and technical systems engaged when a simple vlog (video blog) is uploaded to YouTube. These processes would be further expanded if the creator was to branch out and produce different types of content, collaborate with another creator, cross-publicise, share audiences or co-own channels and profiles.

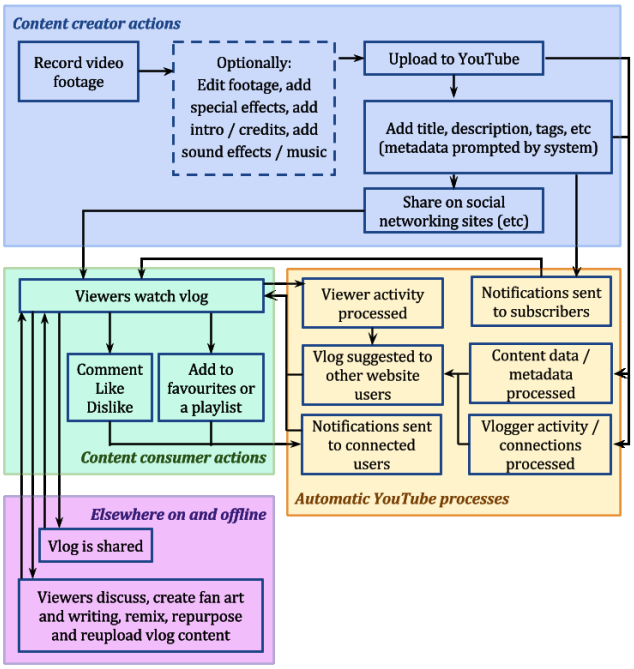


Figure 1: Interconnected social and technical systems necessary for publishing a vlog on YouTube.

Creative media production social machines create an environment in which content creators of all backgrounds and abilities are able to publish outside of the constraints of traditional media channels. These creators are actively vying for attention from content consumers; competing for views, likes and shares on a global scale. How they present themselves to their audience can be critical to their success, but also a ground for playful experimentation.

2.1 Motivations for participation

It is worth noting that there are a variety of motivations or incentives for content creators to participate in these systems.

Some content host sites provide direct financial incentive for popularity (eg. YouTube's Google Adsense). Others facilitate a commission based model, where creators show off their work and take paid requests for custom pieces from the community (eg. DeviantArt). For content creators who publish primarily on such systems, their activity on other systems is usually tied to driving traffic back to the content which makes them money, or entertaining the fanbase from whom they thrive (eg. a creator who publishes sketch comedy on YouTube might use their Twitter account to tell original jokes to maintain interest between video releases).

But for many content creators, the financial rewards from their chosen content host sites might be a convenient side-effect of doing something that they love. Reputation as a creator of high quality content, a talented artist or a particularly funny comedian might be their primary driver. There are also social norms in many communities that affect content creator behaviour. Sometimes creators don't want to be accused of 'pandering' to their audience or losing their artistic integrity, and regulate their behaviour accordingly.

The visibility of quantitative data collected by a content host site — such as how many views a piece of content has, how often they are referred to as a co-creator, or how often a content creator persona responds to viewer comments — may also impact behaviour. Technical factors are often highly conflated with the social norms in a community.

Thus, the core reasons for creating content can affect both the content created and how creators present themselves to their audience in the process.

3. CONSTRUCTED ONLINE IDENTITY

The nature of identity and anonymity in online spaces is well discussed [6, 7, 13, 14]. Humans naturally adjust the way they present themselves according to the context, and different online spaces may afford different levels of flexibility in doing this. Some systems don't require any kind of registration to post content, allowing people to adopt and discard personas as needed, and to create social cues to identify each other that are not designed as part of the system [2]. Entirely different behaviour occurs in systems that strongly encourage or even try to enforce usage of real names. Often it is trivial for people to create multiple accounts under different pseudonyms, but there may be an increased expectation of honesty from other users of the system, which itself affects the culture of communities within.

In many cases the fact that people present themselves differently in different contexts is unconscious; a side effect of their participation in a particular system according to the social norms or technical affordances (eg. their desired username may be unavailable resulting in the forging of new branding around an alternative). In other cases, the creation of alternative personas is engineered and deliberate, either from the outset or as something that has evolved over time. The relationship between individuals and their personas can be one-to-one, many-to-one [4] and one-to-many. In this paper we will focus on the latter case, where one individual may present versions of themselves through multiple personas.

4. STUDY OF CREATOR PROFILES

With the goal of observing some of the different types of identity behaviours exhibited in online media production spaces, we conducted an empirical study of the public profiles of content creator personas, covering a period from 2005 to the present. We examined:

- how they present themselves;
- how they spread their presence across different websites, platforms or communities, and how they distribute their content across their different presences;
- how they present implicit and explicit connections with other versions of themselves, and with other content creators.

We added context to our observations for each creator by taking note of their audience, the type of content they

create, and the capabilities of the platforms on which they publish their content.

The results presented here are based upon a very small (albeit diverse) sample. We later discuss ongoing and future work in improving and extending this study.

4.1 Content creators

Ten content creators were selected from a subset of creators with whose content we have a passing familiarity through encountering it online over the past months to years. This resulted in a broad spectrum of content types (video, animation, music, art, written word) genres (comedy, game commentaries, educational, political), popularity, eminence, and activity levels. We deliberately examined content creator profiles from the perspective of a content consumer, or casual audience member. Thus, for the purposes of this study, we do not have access to deeper insight about the personas beyond what is accessible publicly through the web. Future work that expands on this is discussed later.

4.2 Profiles and personas

Profiles are application- or website-specific and are used to display information about an online persona. An individual may represent versions of themselves through multiple personas, and each persona may have profiles on numerous different websites.

For ten content creators, 93 profiles were discovered. Of these, 23 were YouTube channels, 16 Twitter profiles, 13 Facebook, 9 Vimeo, 7 Tumblr, 6 personal websites, 5 Instagram and 4 Vine, 3 Google Plus, 2 Bandcamp and 2 DeviantArt and 1 each of Patreon, FormSpring, BlipTV, and Newgrounds. Table 1 shows how the profiles are distributed. As we can see, in the domain of creative content production identities are not site- or community-specific. Creators spread their activities across a number of networks in order to shape a richer identity.

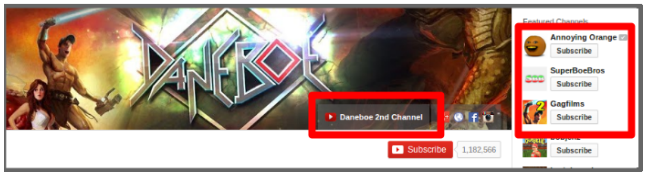
Creator	# profiles	Mean # profiles per site
Dane	18	2.3
Khyan	13	1.9
Bing	13	1.3
Lucas	11	1.4
Bown	9	1.5
Todd	7	1.2
Arin	7	1.0
Suzy	6	1.2
Ciaran	5	1.3
Chloe	4	1.0

Table 1: Distribution of profiles for content creators in the study.

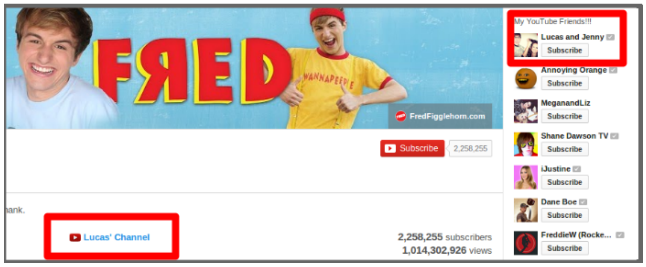
The disproportionate number of YouTube channels is due, in part, to creators having a self-described 'main channel' and 'second channel'. Creators who focus on one type of content (eg. sketch comedy) publish this on their main channel, and use their main channel identity for interactions on the site. On their second channel, for which the audience is typically a more forgiving subset of their main channel audience, they publish things like vlogs about their lives, outtakes or behind-the-scenes footage, or experimental pieces.

Most content creators with second channels post explicit links to them on their main channel, and often publicise them within content metadata or as part of the content directly. In some cases, including those where the connection between two channels is explicit and obvious, the creators present themselves differently towards their audience through second channel content. This varies greatly depending on the type of content produced. In some cases, second channels may be perceived by the audience as more reflective of the creator's 'true' or offline personality; some project themselves as more serious, honest or revealing, and publish more personal content. The significance is that persona variations exist, and creators do not necessarily hide these alternative presentations of themselves from their audience.

Additionally, there are profiles which are *not* directly linked from the (self-identified) 'main' profile, or the links are treated as though they lead to the profile of a different person. Figure 2 shows three screenshots of different YouTube channels showing different ways creators link out to other versions of themselves.



(a) Dane transparently links to three of his channels, two for alternative content types and one for a character he created.



(b) Fred is a character played by Lucas, but the links on Fred's channel treat Lucas as a different person.

(c) Andrew Lemming lists Khyan as "Uncle", although Khyan is the creator of the Andrew Lemming character.

Figure 2: Three creators, who link to different versions of themselves with varying levels of transparency.

Creators also used their profiles to link to shared channels (where either multiple creators post content independently of each other, or creators collaborate to produce joint content, or both), and channels of others with whom they regularly work.

Most profile host sites offer limited options for customisation, and the use of consistent branding was intermittent. The branding mostly took the form of identical or similarly styled display pictures, similarly phrased introductory paragraphs, and similarly styled content.

Some creators have profiles which are distinctly grouped into alternative personas. This was evident from the branding, content and connections between them.

4.3 Connections

How connections to other people were represented varied depending on the technical system. We can differentiate between mutual relationships between accounts (eg. 'friend') and one-directional relationships (eg. 'follower', 'subscriber'). Some systems offer both types of relationship, some one or the other. For YouTube channels, popularity ranged from over 3.5 million subscribers for Dane's character channel *realannoyingorange* to 118 for Bown's secondary *bowntalks* channel.

The importance of these connections varies depending on the system as well as on the attitude of the system user. Mutual connections may initially be presumed to indicate a closer relationship, but this is not always the case. Some systems allow users to accept all friend requests en masse, which they may do to please fans, resulting in a lot of essentially meaningless mutual connections. Outbound one-directional connections come in far smaller numbers, and indicate the content creator is particularly interested in the outputs of the other creators they choose to follow. It is normal for content creators to follow other creators with whom they have collaborated.

The connections are strongly influenced by both community norms and the architecture of the particular website. In addition to inbuilt profile-connecting facilities, most of the websites examined allow enough control over the textual content of a profile that profile owners can manually create links to other documents on the web. Creators may also be able to adapt the content publishing interfaces to add additional connections (eg. adding links out to their Twitter and Facebook accounts in the description of a YouTube video), and often do. These connections necessitate extra effort on the part of the content creator, and tell us more about their relationships with other online accounts. Figure 3 shows different types of connections between profiles and personas for one content creator.

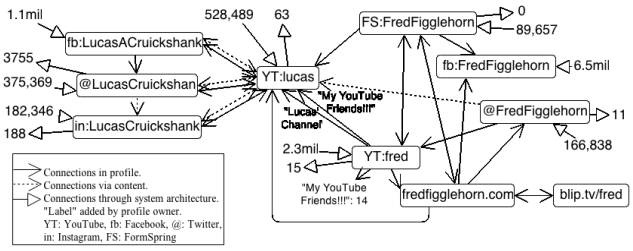


Figure 3: Lucas Cruickshank was an early YouTube success story through his persona Fred Figglehorn. Here, different types of connections between various online accounts belonging to both Lucas and Fred are illustrated.

4.4 Summary

Content creators at all levels of activity have complex and nuanced relationships with the systems they use for publishing and publicising their content. Through manual examination of online profiles it is possible to identify different personas that have been adopted by content creators and the connections between them, and assess the likely explanations behind them. Clarifying these relationships is useful for gaining a deeper understanding of the actions in this space, as well as for creating machine-readable representations of social machine participants. In next section we describe some concepts that may be useful in classifying participants in a social machine based on their identity and connection behaviours.

5. FRAMEWORK FOR ONLINE PERSONAS

Based on the findings previously described, we propose four closely linked but distinct concepts that are useful in a granular discussion of online personas of social machine participants: *roles*, *attribution*, *accountability* and *traceability*. We explain each in the context of creative media production social machines, and give examples of contrasts exhibited in other commonly-discussed social machines. This demonstrates how they can be used as dimensions to assess the nature of individual personas and understand the behaviours of individual social machine participants.

5.1 Roles

The *roles* dimension describes the ease with which participants can change the role they play in a system.

A creative media production social machine contains consumers, commentators and curators, and creators [10]. These roles are interchangeable, and content creators may wish to adopt different personas according to the role they are playing. Moreover, content creators are often multi-talented and they may wish to put on a different face according to the different types of content they publish. How easily this is accomplished, according to the social expectations and technical affordances of a system that is part of a social machine, can impact the behaviour of participants.

Creative media production social machines are systems where there are multiple roles which participants can move fluidly between. In contrast, reCAPTCHA [1] is a system with only one role for all participants, where everyone is equal; and the Obama campaign [12] is a system where multiple roles are available but participants are restricted to playing only one.

5.2 Attribution

The *attribution* dimension allows us to consider whether or not crediting participant contributions is important.

In content creation communities, contributions to media output are directly connected to building reputation, so content creators generally desire to have their name (or that of one of their personas) attached to work they produce. If the publication system does not allow this directly, as is often the case for sites that host collaborative works (a video published on one YouTube channel may contain contributions from several creators, each with their own channels), then creators adapt the system as best they can (the uploader may list links to the channels of all contributors in the video description) [9]. Even when a content host site provides automatic linking to other user profiles — common in remixing communities — this isn't necessarily enough. [11] finds that human-given credit means more, and so free-text fields for content metadata are often used anyway.

Attribution or credit-giving is entirely unimportant for reCAPTCHA, slightly important for Wikipedia (based on the

degree to which participants have invested in the community), and far more important for creative media production.

5.3 Accountability

The *accountability* dimension helps us to decide whether the provenance of the inputs — how much, or what, anyone knows about the contributing participants — make a difference to the function of the social machine.

In many of the oft-discussed social machines, like Wikipedia, Galaxy Zoo, Ushahidi, and the theoretical crime data social machine in [3], accurate data is critical to the usefulness of the output of the system(s). Thus, accountability through identity is important. It is reasonable then to want to regulate participants somehow. But this cannot be universally applicable.

The production of creative content is a domain that exemplifies the need for taking a more flexible approach to identity understanding and management. On one hand, creators wish to be accurately credited for their work and plagiarism may even result in a financial or reputational loss. On the other hand, creators may appear under multiple guises, engage in diverse behaviours and make contradictory statements about their participation in a creative work, all in the name of entertainment. Creators may also engage in some activities under an alternative identity in order to avoid any effect on the reputation of their main persona. These are valid uses of the anonymity and pseudonymity provided by online spaces — a core feature of the World Wide Web — and should be allowed. Such activities won't necessarily result in diminished trust or unstable experiences for other participants. A content consumer may fully enjoy a series of vlogs, unaware that the vlogger is a character and the life events portrayed are entirely fictional, and be none the worse off for it.

5.4 Traceability

The *traceability* dimension relates to the transparency of *connections between* different personas (as opposed to what we know about the background knowledge or experience of a particular persona, as in accountability).

We consider traceability in terms of the settings in which an individual might interact with others. A person participating in a creative media production social machine may adopt a different persona when participating in a scientific discovery social machine, and yet another in a health and well-being social machine. If this individual learns that other participants in the health social machine are aware of their alternate persona as a filmmaker in the creative media production social machine, they may see cause to amend one or both of their personas. If the risk of their multiple identities being 'discovered' is high they may adjust their behaviour accordingly, whether this is ceasing all attempts at 'deception', or taking steps to decrease the overlap of the communities of which they are a part.

Further to this, well known content creators often appear at offline events to meet their fans. Those who star in popular video content are recognised in the street. They are interviewed by journalists and contracted to produce viral adverts by marketing companies. Only with careful control of their online persona can they successfully engage in offline interactions like this. Imagine a content creator who appears in poorly-produced videos as an apparently genuine violent drunk and who sends abusive messages to other website users along this theme to maintain the persona. They may not be considered a candidate for a job in broadcast media, despite the fact that what they have really done is skilfully written and portrayed a believable character.

In this case, high traceability of content creator personas may cause problems if the individuals adopting them are trying to project a particular image. An example in which the traceability of personas was crucial is the DARPA Network Challenge [16], for which participants needed to provide their 'real life' identities to win the cash prizes. Even if they had operated under pseudonyms during the competition, in order to validate their claims they needed to make known these personas and consolidate them with an identity that would allow them to receive the prize money.

6. CONCLUSIONS AND FUTURE WORK

The profiles we have studied are examples of some of the complex identity-related behaviours that occur in social machines. This is very early work, and it is likely that we have only scratched the surface.

6.1 Broader and deeper studies

As well as planning to expand our empirical study to include more profiles, we have identified content creators with whom to conduct semi-structured interviews about their online personas. We expect this to reveal something more than we can understand from their profiles alone. We must acknowledge however that interviewees may not be willing to reveal details of every single version of themselves present on the web. Nonetheless, we can gain a greater understanding of their activities and motivations behind the personas they do choose to reveal.

We will also begin to crawl profiles on all of the major content host sites and automatically gather connection information, and analyse the network that results.

6.2 Ontology for online personas

This study is part of a broader project in which we have been developing a formal ontology for representing activities of content creators. The work presented here has enabled us to gain a better understanding of the subtleties of identity management prior to more extensive formalisation. Designing an ontology that can flexibly represent multiple personas without infringing upon an individual's right to pseudonymity or anonymity will be useful in decentralising social systems. It will also allowing social machines, which may involve alternative versions of the 'same' participants, to interact with one another through a common layer of web standards.

6.3 Conclusion

We have demonstrated through an empirical study that participants in social machines often have complex relationships with their own self-representation, and with their connections to other versions of themselves as well as other individuals in a system. Individuals may have one-to-many or many-to-one relationships with online personas, for a number of different reasons, and with different levels of transparency. We have presented a preliminary framework which allows the degree to which this complexity can affect the overall understanding of a social machine to be assessed along four dimensions: roles, attribution, accountability and traceability.

7. REFERENCES

[1] L. Ahn, M. Blum, N. J. Hopper, and J. Langford. CAPTCHA: Using Hard AI Problems for Security. In E. Biham, editor, *Advances in Cryptology - EUROCRYPT 2003*, pages 294–311. Springer Berlin Heidelberg, 2003.

[2] M. Bernstein, A. Monroy-Hernandez, D. Harry, P. Andre, K. Panovich, and G. Vargas. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 50–57. Association for the Advancement of Artificial Intelligence, 2011.

[3] M. Byrne Evans, K. O'Hara, T. Tiropanis, and C. Webber. Crime Applications and Social Machines: Crowdsourcing Sensitive Data. In *WWW 2013 Companion*, Rio de Janeiro, 2013. ACM.

[4] B. Dalton. Pseudonymity in Social Machines. In *WWW 2013 Companion*, Rio de Janeiro, 2013. ACM.

[5] D. De Roure, C. Hooper, M. Meredith-Lobay, K. Page, S. Tarte, D. Cruickshank, and C. De Roure. Observing Social Machines Part 1: What to Observe? In *CHI'13*, Paris, 2013. ACM.

[6] J. S. Donath. Identity and deception in the virtual community. In P. Kollock and M. Smith, editors, *Communities in cyberspace*, pages 29–59. Routledge, 2002.

[7] H. Halpin. Identity, Reference and Meaning on the Web. In *WWW2006*, Edinburgh, 2006.

[8] J. Hendler and T. Berners-Lee. From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*, 174(2):156–161, Feb. 2009.

[9] K. Luther. Edits & Credits: Exploring Integration and Attribution in Online Creative Collaboration. In *CHI*, pages 2823–2832, 2010.

[10] K. Luther and N. Diakopoulos. Distributed Creativity. In *Supporting creative acts beyond dissemination. Creativity & Cognition*, Washington DC, 2007.

[11] A. Monroy-Hernandez, B. Mako Hill, J. Gonzalez-Rivero, and d. boyd. Computers can't give credit: How automatic attribution falls short in an online remixing community. In *CHI*, Vancouver, 2011. ACM.

[12] K. O'Hara. Social Machine Politics Are Here to Stay. *IEEE Internet Computing*, 17(2), 2013.

[13] S. Rains. The impact of anonymity on perceptions of source credibility and influence in computer-mediated group communication: A test of two competing hypotheses. *Communication Research*, 34(1):100–125, 2007.

[14] Y. Ren, R. Kraut, S. Kiesler, and P. Resnick. Encouraging commitment in online communities. In R. Kraut and P. Resnick, editors, *Building Successful Online Communities: Evidence-Based Social Design*, chapter 3, pages 77–125. MIT Press, 2012.

[15] M. Strohmaier. A Few Thoughts on Engineering Social Machines. In *WWW 2013 Companion*, Rio de Janeiro, 2013. ACM.

[16] J. C. Tang, M. Cebrian, N. A. Giacobe, H.-W. Kim, T. Kim, and D. B. Wickert. Reflecting on the DARPA Red Balloon Challenge. *Commun. ACM*, 54(4):78–85, 2011.