Surveil and Calculate

Katharina Spiel

Faculty of Media
Bauhaus-Universität Weimar
Bauhausstraße11
99423 Weimar
katharina.spiel@uni-weimar.de

Abstract

Digital surveillance techniques are currently based on statistical means of extracting abnormal behaviour, situation or persons. The implications of these techniques go beyond the digital realm and not only affect the physical, but are reshaped by their own outcomes. This is not only true in the context of surveillance, but also digital communication and interaction with virtual agents. Developers are encouraged to take this into account when creating software monitoring humans including their means of interaction with the software.

1 The Digital Panopticon

In 1999 Andy and Lana Wachowski questioned the concepts of reality and virtuality in *The Matrix*[1]. Their metaphor of a human life encapsulated in an artificial virtual reality created by machines culminates in the very term used to describe it. Whereas matrices in mathematics denote containers for numeric information, the Wachowski Matrix is the container for the majority of human lives that are perceptually integrated and, hence, subject to the mathematical-numerical logic in which they are embedded.

The architecture within the movie in which humans are arranged outside of the matrix is an inverted version of Jeremy Bentham's panopticon (see Bent95) of 1785. Whereas in Bentham's prison the viewpoint is central to see all of the converging cells at once, outside the matrix the cells diverge (cf. Figure 1). It is not the physical architecture which follows rules of panopticism, but rather the emulated social architecture which has to be constantly maintained

[1] *The Matrix,* 1999, motion picture, Warner Bros. et.al., Australia and USA.

and surveilled by the machines who maintain the matrix as a means of survival. „Visibility is a trap.“ (p. 251, Fouc76)

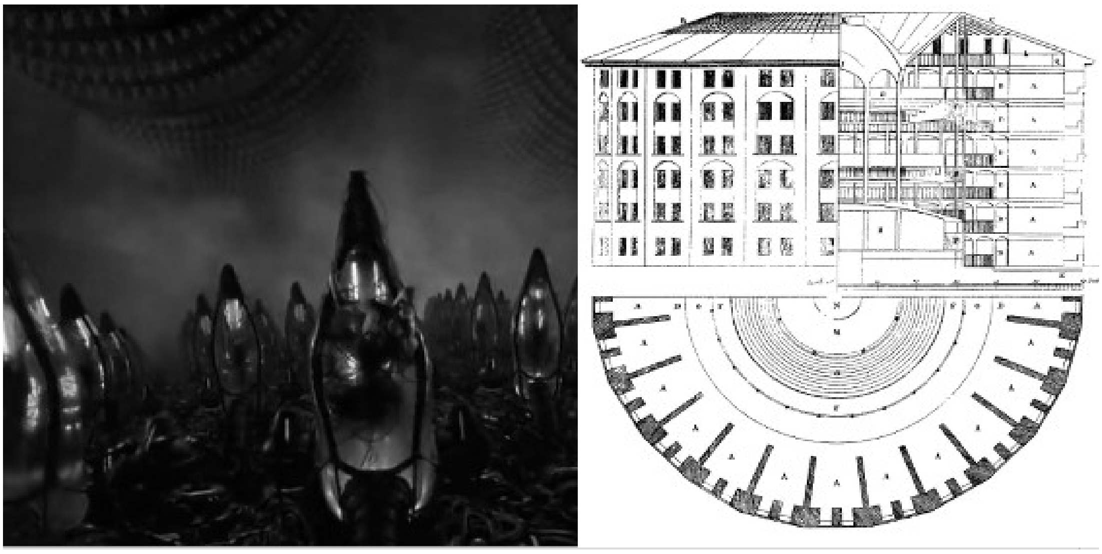

Fig. 1: Left: Physical Architecture outside the Matrix (source *The Matrix*);
Right: Jeremy Bentham's Prison (drawn by Willey Revely, 1791)

Where the movie exaggerates the concept of a digital matrix as a futuristic secondary mode of existence, the analogy holds for digital communication. „The social technologies we see in use today are fundamentally panoptical – the architecture of participation is inherently an architecture of surveillance.“ (Ross09) Where communication becomes digital, the panopticon becomes digital too. However, there are key differences.

In a physical world, a person can be seen as a singular entity performing different roles (see Jone00). In a digital world, on the other hand, a single identity can be used by multiple people (see Lunc09 on the single identity of multiple hackers) and multiple identities can be held by a single (see DiMi07 describing personae in different social online platforms used by the same physical person), where none of them have to coincide with that person’s appearance or identity in the physical world (see Dona96 on so called deceptive online identities). In order to cross-reference the physical person with their online personae, several methods have arisen, such as identification algorithms (see Ma07 for a thorough analysis of methods) and strategies to ensure that users use their birth names within online services (see Face14, „Facebook is a community where people use their real identities.“).

In *The Matrix*, Neo can redefine his own identity and the machines try to identify him within the matrix, whenever he enters it (after having been emancipated to do so by choice). However, whenever he enters the matrix, his experiences in the matrix shape his experiences and evaluations outside and vice versa. Hence, there are relationships between the outside world and the matrix for Neo as well as interdependencies between the digital identities of people and the identities used for their physical persons. Due to these interdependencies it is important to look at the effects of statistical analysis of large scale data, new surveillance techniques emerging from it as well as how normality is shaped under these conditions.
The field of surveillance is then of special interest, since „[the] digitised fabric is characterised by a pervasive field of information, code, and signifiers that increasingly construct the social environments that mobile bodies pass through and negotiate. Near-constant surveillance of the person, both in public life and in private affairs has blurred boundaries between what is external and what is internal, between outer and inner freedom." (p. 347, King08). This means, that identities and their presentations blur as well. However, in software which statistically analyses its users, this blurring is not accounted for, resulting in a definition of normality which constantly becomes narrower.

2 Numbers

2.1 Statistics

Stochastic methods are used in machine learning, natural language processing and data analysis both big and small. Statistics is „the business of interpreting observational data, so as to obtain a better understanding of the real world“ (p. 69 Fish55). It is not a tool to measure an event, but rather enables observers to make data-driven deductions helping with the analysis of their subject of interest. „[We are] not speaking of the *final* acceptance or rejection of a *scientific* hypothesis on the basis of statistical analysis. We speak of accepting or rejecting a hypothesis with a "greater or less degree of confidence". Further, we [are] very far from suggesting that statistical methods should force an irreversible acceptance procedure upon the experimenter“ (p. 206 Pear55).

While stochastic calculations are often applied in rhetorics to suggest authority and scientific proof, the only thing that statistics say is whether a result comes from an analysed effect or whether it could also just been created by random occurrences with more or less confidence.

However, to create the experimental setup and hence analyse data, categories first have to be created. Usually, this is done by the experimenter themselves. So, "the whole purpose of a series of events relies on the prinicples of order that the observer pulls over it. This is of essential meaning for our perception of reality" (p. 72 Watz76). Due to this, exact precision cannot be expected when data is being analysed by statistical means. Their „logical differences (…) seem (…) so wide that the analogy between them is not helpful, and the identification of operation is decidedly misleading" (p. 70 Fish55).

2.2 Categories

Before statistical methods can be used, categories for analysis have to be defined, which already shape the outcome of the calculations. Hence, it cannot be assumed that they represent absolute truth and thus have to be approached critically. In the context of surveillance and automated virtual communication, likelihoods might be the most precise tool with which to create software, however, the categories and the data on which they operate are often not critically examined.

The concept of categories and an order of things is essential to distinguish between normal, expected behaviour and abnormal, unexpected behaviour or just differentiate between match and mismatch. Where statistics work with confidence levels to make a statement that includes a certain amount of uncertainty, categories are discrete entities into which things are sorted. While new categories sometimes emerge when necessary, things find themselves in a binary state concerning a specific category: they are either part of it or not.

When things are assigned to a category by stochastic means, they are assigned with a certain confidence. This is not a binary decision as such, but algorithms often take the confidence level and as soon as it is above a certain threshold, assign an item to a category, forgetting the possibility that it might not be part of that category. For this to happen, it does not matter whether the categories are predefined (in supervised learning algorithms) or not (in unsupervised learning algorithms). The distinction between those two types of algorithms lies in a tiny difference in their assumptions. Both assume that the data can be sorted into categories. In supervised learning algorithms, there is the additional assumption that the data can only be sorted into categories which were defined beforehand.

Categories in unsupervised learning are supposed to emerge from within the data and can change when the data changes. Whether the number of categories has to be given beforehand or not depends on the algorithm used to categorise the data. Since categories then do not necessarily hold any semantic meaning, they are often referred to as clusters (see Witt05, p.81).
Supervised learning is used, whenever a category is known beforehand and there is labelled data to train on. One common method is Bayesian statistical modelling (see ib. pp. 88), often found in surveillance algorithms (cf. Meng11 or Burk13 as examples) but also in approaches to create augmented realities (cf. Coir07 or Ong07 as examples). As in statistics these models work with heuristics which shape the outcome before any data has been analysed.

3 Knowledge

Fouc81 argues that there is not only one global truth and that knowledge is created by a discourse eventually leading to what can be mutually agreed upon as part of knowledge[2]. In the case of stochastically driven implementations of digital surveillance and communication technologies, this procedure becomes obvious. Since the target can only be analysed along predefined or emergent categories, some of which hold no semantic meaning, knowledge about the target can only be acquired within the parameters the categories themselves set.
For most of society, knowledge about this software can only be discovered implicitly via a user's assumption on how it is constructed, which does not necessarily relate to the actual instructions and rules followed by the program. This also holds true for the paradigms behind the creation of software. Usually, not even the programmers themselves fully understand the entire code path from higher level language to machine code. Not only „graphic user interfaces [...] hide a whole machine from its users" (p.3 Kitt95). Assumptions, hence, are a necessary part of processes dealing with software.
When users interact with programs, they do so either knowingly, like in contexts of augmented or virtual reality, unknowingly, like in contexts of surveillance or wih the expectation of a different functionality, like when communicating online with an illusion of privacy. However, these cases seldom show up in their pure form, so while knowingly interacting, there are still elements of unknowing interaction.

[2] The work has later been expanded and made concrete for sciences by Lato00.

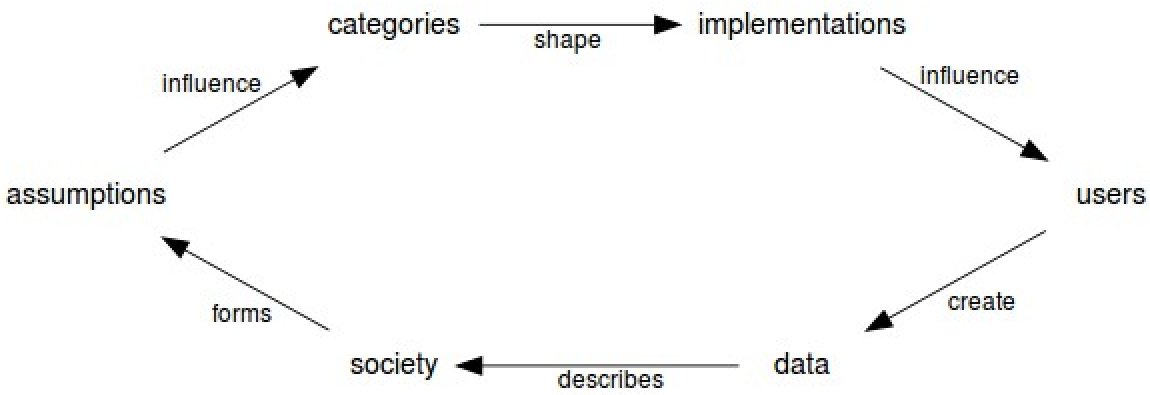

Fig. 2: The Interference Cycle of the Programming Pipeline. Other relationships between the individual steps exist, but have been omitted to demonstrate the loop.

When software is in use, an interference cycle can be established (see Fig. 2). As shown beforehand, the developers' assumptions influence the categories which have to be established in order for the software to perform its calculations and hence shape the implementation. When interacting with the specific implementation, users become influenced by that software via their modes of interaction. The data created by this interaction describes societies to which the users belong. Ideally, these societies are closely watched by software developers, because their future users will come from them. Hence, the information about these societies again forms the developers' assumptions, leading to categories that influence implementations and so on around the cycle. This happens in all forms of interaction: when users interact with software knowingly, unknowingly or under false pretenses. They are embedded in a larger cycle of interferences that creates knowledge about its other components.

4 Normalisation

Taking the example of surveillance technologies, Fouc76 points out how modern surveillance is aimed at disciplining the surveilled target. The goal is for the target to leave the realms of the Abnormal[3] and to reintegrate into the normalised bounds of society. This assumes that the target has once been within these bounds.

Digital technologies of surveillance operate via statistical means when categorising elements of society, situations or behaviour into normal or abnormal, civilian or terrorist. When learning

[3] An extensive study on the Abnormal can be found in Fouc74.

from available data, this happens under the assumption that the observed data can be divided into normal and abnormal.

Counter strategies against known applications of surveillance technologies discipline those employing them indirectly. The goal is to avoid individual categorisation as abnormal in order to appear normal. Alongside this, the distribution of data points or individuals in society also changes. Understanding this in the form of a normal distribution, variance within society is affected both positively and negatively. If a society consists of many different identities in a variety of representations, the data points which would be identified as abnormal move toward the middle, again limiting the range of what is normal (see Fig. 3). Since the core assumption, that data points have to be sorted into normal and abnormal still holds[4], new software will work on the new data created and the process starts anew with new counter strategies.

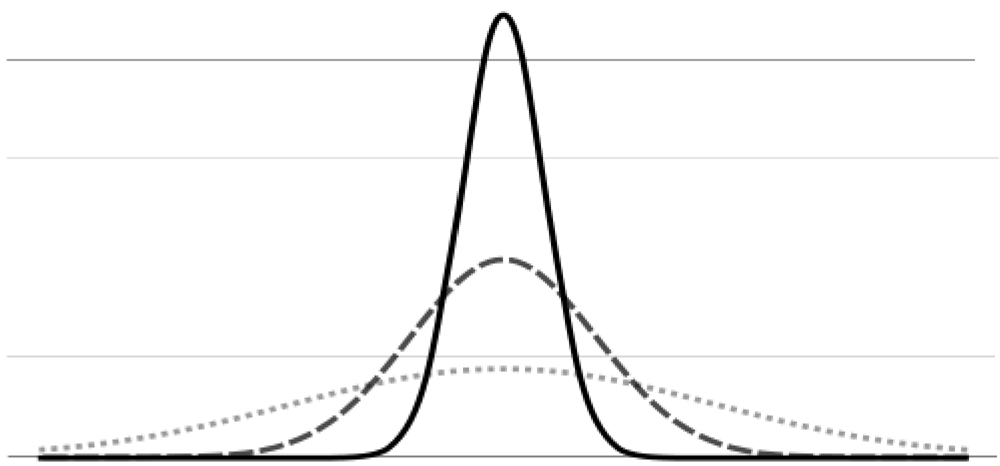

Fig. 3: Shaping of the Normal when Affected by Normal Distributions

Essentially, what accounts for the bell curve, also accounts for the development of society under algorithmic surveillance: The smaller the variance, the higher and narrower the normal distribution will be.

[4]The assumption, that these categories hold is fuelled by the fact, that fear of terrorists is being communicated thoroughly in North-American culture (cf. Myth06).

5 Communication

Digital communication follows mathematical rules as does communication whenever there is at least one virtual agent participating. This will be demonstrated by two examples: the online platform Twitter[5] and virtual guides in a museum.

5.1 Twitter

Twitter is well known for limiting its users to an extremely short style of communication with a maximum of 140 characters allowed per message. This is true for both public messages (tweets) and direct messages – their equivalent of one-to-one communication. The latter type of messages is defined by Twit14 as private, however, users often do not consider that the messages are still stored on a third-party server and hence also available to individuals who were not intended parties of the conversation. Additionally it can happen, that the data which should only be visible to the account holders is given to other parties, e.g. governments[6] or advertiser. Users can partly circumvent formal limitations, for example when posting multiple connected messages in a row. They cannot, however, change systematic limitations, like where their data is stored and with whom the company shares it as soon as they use the service.

Users on Twitter are already accustomed to the limitation regarding message length. It is not expected that this limitation will change, because it is a distinguishing attribute of the service. Another aspect of Twitter is that most communication on it is intended to be public. Users either publish accordingly or do not realise the dimension in which they are posting information. The first group of users can be seen as somewhat knowledgeable whereas the second group uses the service without being aware of all consequences; their interaction happens under different terms than expected.

When knowledgable users publish only those things which they consciously want to have public, they discipline themselves. The second group of users is only potentially disciplined, e.g. when a user has a bad experience with messages they did not necessarily expect to make available or is being chastised (often publicly) for being too open about their private life. Both cases influence how messages look and normalise interaction on Twitter.

[5]http://www.twitter.com

[6]Twitter publishes statistics on how often this is the case under https://transparency.twitter.com/.

5.2 Virtual Guides

In order to establish communication between humans and virtual guides, e.g. in a museum (see as selected examples Kopp05 or Swar10), these virtual guides have to be able to detect situations of interaction and react accordingly. This can be the user asking a question about a certain object or just where the nearest bathroom might be. However, the virtual guide is not likely to respond appropriately to situations it cannot categorise within a known set. New interaction techniques evolve when virtual guides and users come together and these situations are recorded to be included into the next iteration of the guides.

The interference cycle then becomes similar to a hermeneutic spiral, where change can occur at each step. Assumptions about possible user input influence the categories of situations and the reactions which can be implemented in future development cycles. The options which users discover for interaction and those which stay undetected can be recorded and then create data. Furthermore, users' attempts to interact with a guide which are not met with an appropriate response, can enrich the data. This data then describes desired modes of interaction and how they are performed as well as how users, as part of societies, expect the guide to act. The assumptions for the next iteration of the virtual guide are then formed from the previously acquired data. In order to satisfy the needs of users, a virtual guide should be conceptualised in such a way that it is iteratively under constant development to address societal changes, which also might emerge due to the interaction with the virtual guide.

6 Variance

In *The Matrix* Neo consciously decides when he enters. Eventually he changes the matrix and his physical world with it. Digital communication is not always a conscious decision. The differences between physical and digital realities are not always in the foreground. Combined with the scale on which surveillance and communication technologies are now being used, users are within a digital panopticon, even if they do not realise this. The effect on societies has yet to be fully established.

This work set out to investigate how the definition of normal is established when statistical parameters are used, such as in a digital panopticon. Furthermore, it has been shown that digital and physical communication have influence beyond their own realm.

The calculated normal has effects on parameters inside as well as outside of statistics and influences digital and physical realities alike. Developers should consider, that their methods are shaped by the users or targets as much as the latter are influenced by the software they use or which is used on them. Developers are encouraged to develop software monitoring users incrementally and critically reflect their position in the interference cycle.

In a statistical context, surveillance of users and data is not always done by meaningful extraction and observation of individual cases, but rather by detecting and focusing on outliers. The application of statistical principles to surveillance and communication algorithms makes everything and everyone a data point and hence a potential object of interest.

Literature

[Bent95] Bentham, Jeremy: The Panopticon Writings. Ed. Miran Bozovic. Verso, London 1995, p. 29-95

[Burk13] Burkom, Howard, Elbert, Yevgeniy, Ramac-Thomas, Liane, Cuellar, Christopher, and, Hung, Vivian: Refinement of a Population-Based Bayesian Network for Fusion of Health Surveillance Data. In: Online Journal of Public Health Informatics 5.1, OJPHI, Chicago, IL, 2013

[Coir07] Coiras E., Mignotte P.-Y., Petillot Y., Bell J., and Lebart K.: Supervised Target Detection and Classification by Training on Augmented Reality Data. In: Radar, Sonar & Navigation, IET 1.1, IEEE, 2007, p. 83-90

[DiMi07] DiMicco, Joan M., and Millen, David R.: Identity Management: Multiple Presentations of Self in Facebook. In: Proceedings of the 2007 International ACM Conference on Supporting Group Work. ACM, Sanibel Island, FL, 2007, p. 383-386

[Dona96] Donath, Judith S.: Identity and Deception in the Virtual Community. In: Kollock Peter, and Smith, Mark: Communities in Cyberspace. Taylor and Francis, UK, 1996, p. 29-59

[Face14] Facebook: Facebooks Name Policy, 2014. Available online at https://www.facebook.com/help/292517374180078, last checked March 11, 2014

[Fish55] Fisher, Ronald A.: Statistical Methods and Scientific Induction. In: Journal of the Royal Statistical Society 17.1, The Royal Statistical Society, London, UK, 1955, p. 69-78

[Fouc74] Foucault, Michel: Abnormal: Lectures at the Collège de France, 1974-1975 Vol 2. Macmillan, London, UK, 2007

[Fouc76] Foucault, Michel: Überwachen und Strafen – Die Geburt des Gefängnisses. Suhrkamp, Frankfurt am Main, 1976

[Fouc81] Foucault, Michel: Archäologie des Wissens. Transl. By Ulrich Koppen, Suhrkamp, Frankfurt am Main 1995 (1981)

[Jone00] Jones, Susan R., and McEwen, Marylu K.: A Conceptual Model of Multiple Dimensions of Identity. In: Journal of College Student Development 41.4. American College Personell Association, Nashville, TN, 2000, p. 405-414

[King08] Dennis, Kingsley: Viewpoint: Keeping a close watch-the rise of self-surveillance and the threat of digital exposure. In: The Sociological Review 56.3, Wiley, Hoboken, NJ, 2008, p. 347-357

[Kitt95] Kittler, Friedrich: There is No Software. In: ctheory 10.18, 1995 available online at http://www.realtechsupport.org/UB/MC/Kittler_NoSoftware_1995.pdf, last checked February 21, 2014

[Kopp05] Kopp, Stefan, Gesellensetter, Lars, Krämer, Nicole C., and Wachsmuth, Ipke: A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In: Intelligent Virtual Agents, Springer, Berlin Heidelberg, 2005, p. 329-343

[Lato00] Latour, Bruno: Die Hoffnung der Pandora: Untersuchungen zur Wirklichkeit der Wissenschaften. Suhrkamp, Frankfurt am Main 2000

[Lunc09] Lunceford, Brett: Building Hacker Collective Identity One Text Phile at a Time: Reading Phrack. In: Media History Monographs 11.2, Elon University, Elon, NC, 2009, p. 1-26

[Ma07] Ma, Meng, and Agarwal, Ritu: Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. In: Information Systems Research 18.1. Informs, Catonsville, MD, 2007, p. 42-67

[Meng11] Meng, Hongying, Appiah, Kofi, Hunter, Andrew, and Dickinson, Patrick: FPGA Implementation of Naive Bayes Classifier for Visual Object Recognition. In: 2011 IEEE Computer Societey Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2011, p.123-128

[Myth06] Mythen, Gabe, and Walklate, Sandra: Communicating the Terrorist Risk: Harnessing a Culture of Fear?. In Crime, Media, Culture 2.2. Sage, New York, NY, 2006, p.123-142

[Ong07] Ong, Soh-Khim, Chong J.W.S., and Andrey, YC Nee: Methodologies for Immersive Robot Programming in an Augmented Reality Environment. In: Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia. ACM, Ney York, NY, 2006, p. 237-244

[Pear55] Person Egon S.: Statistical Concepts in the Relation to Reality. In: Journal of the Royal Statistical Society, Series B 17, The Royal Statistical Society, London, UK, 1955, p. 204-207

[Ross09] Ross, Joshua-Michéle: The Digital Panopticon. May 20, 2009. Availabe online at http://radar.oreilly.com/2009/05/the-digital-panopticon.html, last checked March 10, 2014

[Swar10] Swartout, William, et.al.: Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. In: Intelligent Virtual Agents, Springer, Berlin Heidelberg, 2010, p. 286-300

[Twit14] Twitter Support: Types of Tweets and Where They Appear, 2014. Available online at https://support.twitter.com/articles/119138-types-of-tweets-and-where-they-appear, last checked March 12, 2014

[Watz76] Watzlawick, Paul: Wie wirklich ist die Wirklichkeit? – Wahn, Täuschung, Verstehen. Piper, München, 1976/2013

[Witt05] Witten, Ian H., and Eibe Frank: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington, MA, 2005