

## 9. Unknowing Algorithms

On Transparency of Unopenable Black Boxes

*Johannes Pafsmann & Asher Boersma*

### Algorithmic Black Boxes as a Challenge for Media Studies

The primary source for the suspicion with which the rise of and subsequent dependency on software as research instrument in the humanities is met, is that one does not know what *the machine* does. In many cases ‘machine’ means algorithm. *Algorithmic black boxes* have become so widespread that this objection could already be voiced as soon as a researcher uses Google. In *digital methods* and beyond, there is a dominant tendency for research processes to be dependent upon algorithmic black boxes, which even theoretically cannot be ‘opened’ (Bucher 2012). Kate Crawford speaks in this context of the ‘disappointingly limited calls for algorithmic “transparency”, which seem doomed to fail’ (2016: 11).

The dependency on algorithmic black boxes has been addressed as a problem for research practices by Bernhard Rieder and Theo Röhle (2012). They have called ‘black-boxing’ one of the major challenges for digital methods, and continue their pursuit for a solution along the same lines in this volume. They delineate this technical black-boxing as a matter of accessibility (such as in the case of ‘the’ Google algorithm or countless other proprietary algorithms) and code literacy (cf. *ibid.*: 76), but also as not-understandable on a ‘more abstract’ level, as ‘the results they produce cannot be easily mapped back to the algorithms and the data they process’ (*ibid.*). Still, Rieder and Röhle propose this should not keep us from using them, as there is a workaround to this, which is ‘to use different tools from the same category whenever possible in order to avoid limiting ourselves to a specific perspective’ (*ibid.*: 77). Different algorithms would bring different aspects of a data set to the fore when one experiments with them, switches between different ones, etc.. Thus what Rieder and Röhle have proposed – and continue to seek for in this volume with their focus on the ‘bizarre amount of knowledge we have stuffed into our tools’ (Rieder & Röhle in this book) – are ways to minimise the size of black boxes by enlightening formerly black parts.

With this article however, we would like to draw attention to an approach from a different direction. Instead of focusing on how to gain positive

knowledge in order to make these black boxes ‘more transparent’, we would like to outline a concept of transparency that is not so much concerned with positive knowledge, but that deals with skills which help dealing with those parts of an artefact that one still does not know. In that sense our proposal comes in after Rieder and Röhle have left the scene successfully: we ask how to behave towards what remains black after all. We consider this important, as we assume that the minimising strategy will be decreasingly helpful in the future. The conjuncture of algorithmic black boxes is so huge and encompassing that questions of how to minimise the unknown will increasingly be replaced by questions of how to behave towards the unknowable. What appears debatable may thus be not how to make a given black box more transparent so much as the concept of transparency itself.

What we found is that different authors have given significantly different meanings to the term transparency. We would like to differentiate between two notions here. On the one hand, there is what we call *formalized transparency*, which largely tries to obtain ‘more positive knowledge’ on ‘the content’ of a black box. On the other hand, we see *practical transparency*, which does not try to open black boxes, but to develop skills without raising the issue of openness. These two concepts of transparency do not exclude each other. Rather, they outline two different sets of practices dealing with black boxes which can complement each other.

### **Don’t Open Every Black Box!**

When Sociology of Science and the movement that was later termed *Laboratory Studies* adopted the term *black box* in the 1970s, its function was somewhat different from today. Richard D. Whitley stated at the time: ‘The view of scientific knowledge maintained by much of the sociology of science had led to an ideology of “black boxism” which restricts research to the study of currently observable inputs to, and outputs from, a system. Any study of the internal processes, which may be unobservable at the moment, is declared taboo.’ (1972: 63)

The problem that he points out here is not that in research practice there are too many intransparent material black boxes, but that the ‘site of scientific action offers a unique opportunity to investigate the process of knowledge production, which continues to be a “black box” to social studies of science’, as Karin Knorr Cetina put it ten years later (1982: 102).

Latour and Woolgar also talk critically about black boxing, stating that with the help of ‘money, authority, confidence’ certain kinds of knowledge

are prevented from being questioned, making it almost improbable to raise alternatives (1979: 242). This is quite similar to the concept of black boxing that Gregory Bateson uses some years earlier:

A 'black box' is a conventional agreement between scientists to stop trying to explain things at a certain point [...] It's a word that comes from the engineers. When they draw a diagram of a complicated machine, they use a sort of shorthand. Instead of drawing all the details they put a box to stand for a whole bunch of parts and label the box with what that bunch of parts is supposed to do. (Bateson 1972: 39)

What all these black boxes (Whitley; Knorr Cetina; Latour & Woolgar; Bateson) have in common is the normative implication that they should actually be opened. In her later work, however, Knorr Cetina stresses that certain black boxes in scientific research processes cannot and should not be opened. She mentions the example of heart surgeons who want to be present when a donor heart is removed from a body, to see and feel the organ they will have to transplant later. Nobody can explicitly say what the use of this is, or what kind of information the surgeon gets or processes when he or she is present at the removal, but everybody agrees this is necessary:

The scientist's body as an information-processing tool is a black-boxed instrument. The absence of discourse concerning embodied behavior corresponds to the use of embodied information processing as a substitute for conscious reflection and communication. The acting body works best when it is a silent part of the empirical machinery of research. (Knorr Cetina 1999: 99)

This means that in research practices we find black boxes that should not be opened. These black boxes remain intransparent in a conventional, formalized sense, but that is in no way problematic for the scientist in the practice of research. She (and none of her colleagues or the ethnographers researching laboratory life) sees, hears, etc. what is happening, but this produces no deficit for her. Quite on the contrary, an explication of 'what is actually happening' may deprive her of these kinds of 'knowledge'. This is quite similar to when somebody starts thinking about their PIN code when typing it at the supermarket register: the fingers 'know' the code, but as soon as 'the brain' tries to think about it explicitly, the fingers will 'forget' it. This raises the question whether the calls for transparency in discourses around algorithms actually have to point to what Knorr Cetina

called 'conscious reflection and communication' when conscious reflection is impossible anyway, or if here transparency may mean to embody the functioning and dysfunction of a certain artefact: if we cannot 'know' (in an explicit sense) what algorithms do and which inaccuracies they have, can we at least embody them to such a degree that we know when to rely on their results and when to become distrustful? This would mean that they become transparent, not insofar as 'one sees what is happening', but rather in such a way that they withdraw in practice.

### **Practical Transparency: Knowing When and How to Be Careful**

The notion that we call practical transparency has already been applied in the realm of digital media research. Susan Star, Geoffrey Bowker and Laura Neumann write that at the 'individual level of scale', transparency means a user 'does not have to be bothered with the underlying machinery or software. In this sense, an automobile is transparent when a driver sits down, turns the key, and drives off, all without the foggiest notion of how internal combustion works' (2003: 242). This means, and this is crucial, that black boxing is not an obstacle for transparency here, but the primary condition for its very possibility.

Applying this notion of transparency to algorithmic black boxes would only be half the solution as it does not provide means to position oneself critically towards what is inside the black box. This concept however has a long tradition in the philosophy of technology, and has actually always been an ideal of media and media practices. It is here that we find an understanding of transparency and its limits that allows for the critical position we seek.

Most prominently, we find this idea in Marx's *Maschinenfragment*. Originally however – although Marx does not mention this – the idea comes from Hegel. Loosely speaking, Hegel differentiates *Maschine* (machine) and *Werkzeug* (almost translatable as *tool*). The tool follows the hand, whereas the hand has to follow the machine. The tool mediates between man and the environment, as an inert but still rather passive thing in the producer's hand – the German word *Werkzeug* expresses this hierarchical relation, *Zeug* is the stuff, the unimportant, the heteronomous. That the tool is inert means that it forces its user to discipline himself, as Axel Honneth comments, with the result that on the one hand, the user 'transforms himself into a thing'. But at the same time he experiences that in this self-disciplining, the subjective *Geist* acquires the ability to realize itself in the product of

tool-using work (cf. Honneth 2014 [1994]: 61). This would mean that for the ideally disciplined Hegelian tool-user the tool is practically transparent.

The machine on the other hand requires a different type of disciplining. At first it seems that the ‘user’ lets the machine work for himself, ‘but his own activity thereby becomes more formalised. His dull work constricts him to a single point [...] shrinking [his] skills” (Hegel 1983 [1805/6]: 139). At the end of this process, even this formal work is *aufgehoben* (sublated, abolished, suspended, superseded, but also preserved). The machine working for you ultimately inverts your relation to it: you work for the machine, doing only machine work. Hegel considers machine usage as a fraud against nature, which in turn takes revenge: the more man subjugates nature using machines, the lower he himself sinks (cf. Hegel 1974 [1803/4]: 332). The tool thus appears in that respect as the opposite of the machine, as it follows the intentional activity of its users and can be made transparent as they learn to master it with certain skills, whereas the machine on the other hand makes its users transparent by mastering them.

Our differentiation between Hegel’s concepts of *Maschine* and *Werkzeug* is, as mentioned, relatively rough. The German philosopher Hans-Christoph Schmidt am Busch is a bit more precise here; he differentiates between tool as a means of labour (‘Werkzeug als Arbeitsmittel’) and tool as a thing (‘Werkzeug als Ding’), and the actual tool is always both (2002: 48ff). Similar to the algorithm that allows you to make certain parts practically transparent (one can ‘know’ what they do) and others not (one can ‘know’ their inaccuracies), the Hegelian tool also has a dual character as *means* and *thing* at the same time. The ideal however is a practical transparency that can be achieved through acquisition of skills.

This process of appropriating a technical artefact in such a way that it becomes transparent through acquired skills may actually be one of Maurice Merleau-Ponty’s most prominent considerations. Certain artefacts can become transparent in such a way that they become part of one’s body schema, they are incorporated or, as he writes, *embodied*. Merleau-Ponty names several examples of embodiment, one of them being the blind man’s stick, which withdraws when used. At the end of a learning process, for its user, the stick is no longer an object, as he does not perceive it as anything distinct from his body: the stick’s ending is the beginning of its user’s perceptual field, and the stick’s measurable length no longer matters (1962

1 This quote is from Leo Rauch’s translation and refers to *Jenaer Realphilosophie* (1805-06), chapter II ‘Wirklicher Geist’, subchapter b ‘Vertrag’.

[1945]: 167)<sup>2</sup>. The world of tactually perceived objects then does not begin with his skin, but with his stick's ending. Thus he does not *interpret* the pressures the stick makes towards the hand. On the contrary, embodiment means that explicit interpretation is no longer necessary, just like it is not necessary to interpret contact with one's skin (cf. *ibid.*: 178).

For Merleau-Ponty, the same principle works for other artefacts like typewriters and pipe organs. Two examples that he gives are particularly interesting for us. The first is a woman with a plume hat, who can keep a safe distance between the feather in her hat and everything that may break it off without calculating the distance between the top of her feather and the respective objects. The same goes for the second example, a car that one wants to drive through a tunnel. Car and hat do not function as objects with measurable volumes, but have become demands for a certain amount of free space (*ibid.*: 167)<sup>3</sup>.

The feather and the stick are both 'extensions of man', but they bring quite different aspects of these extensions to the fore. The stick appears as a transparent mediator to perceive *and* act upon the environment. The feather on the other hand does not appear as a mediator here (which does not mean it could not be used as such, but it's just not as suited for that kind of task as the stick is). It does not so much produce practical transparency, but forces you to learn when and how to be careful. As already mentioned, for practices involving algorithmic black boxes, both appear important: to embody what they can do, that is Merleau-Ponty's stick (which corresponds to Hegel's notion of *Werkzeug als Arbeitsmittel*) and to embody its inaccuracies as well, which is represented by Merleau-Ponty's feather (and corresponds to Hegel's notion of *Werkzeug als Ding*).

Thus the feather highlights the known unknown that you need to embody in a particular way if you want to produce knowledge with algorithms, whereas the blind man's stick highlights the embodied known or not explicitly known known. When both kinds of knowledge about algorithms, the positive 'stick knowledge' and the negative 'feather knowledge', are taken together, it appears possible to act towards unopenable black boxes.

2 'Le bâton de l'aveugle a cessé d'être un objet pour lui, il n'est plus perçu pour lui-même, son extrémité s'est transformée en zone sensible.'

3 'Le chapeau et l'automobile ont cessé d'être des objets dont la grandeur et le volume se détermineraient par comparaison avec les autres objets. Ils sont devenus des puissances volumineuses, l'exigence d'un certain espace libre.'

In one of our earlier articles, a case was presented where two researchers used Gephi algorithms to visualize Twitter data (Paßmann 2013). There, exactly, this difference became apparent: the researchers differentiated between certain activities of an algorithm that they would undoubtedly consider solid results and for which they would take responsibility – this was for example the case when the distance between two nodes in a network visualization was 20 cm. On the other hand, there were results on the map which they did not ‘dare to say anything about’ (*ibid.*) – in this case when the distance between two nodes was only two cm. The latter would be this ‘feather knowledge’ telling its users when and how to be careful (demanding a certain amount of free space, i.e. the known unknown), whereas the 20 cm, after sufficient experience, would be in the realm of the ‘stick knowledge’. When approached through the lens of formalized transparency, the whole algorithm would appear as a black box here, which with the help of Rieder and Röhle can be cut down to a significantly smaller size. In terms of practical transparency, the remaining part that we formally cannot know can be practically transparent either through embodiment, as with the stick, or through a carefully paced out unknowing, as with the feather.

One result of such ‘feather knowledge’ would then be to realize which other sources, external to software and database, are necessary to work around the known unknowns, like ethnographic data for example. Regularly these other sources will be some kind of everyday knowledge, which has not been optimized as much as, for example, knowledge about software and databases. That means that, at best, practical transparency turns unknown unknowns in to known unknowns. Finding ways to deal with these new known unknowns needs more attention, we would argue, than the inner workings of black boxes. We know they will be increasingly unopenable.

## References

- Bateson, Gregory. 1972. *Steps to an Ecology of Mind*. Chicago: University of Chicago Press.
- Bucher, Taina. 2012. “Want to be on the top? Algorithmic power and the threat of invisibility on Facebook.” *New Media & Society* 14 (7): 1164-1180.
- Crawford, Kate. 2016. “Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics.” *Science, Technology & Human Values* 41 (1): 77-92.
- Hegel, Georg Wilhelm Friedrich. 1974 [1803/1804]. “Anhang zur Jenaer Realphilosophie, Ausarbeitung zur Geistesphilosophie von 1803/4.” In *Frühe politische Systeme*, ed. Gerhard Göhler, 293-335. Frankfurt a.M./Berlin/Wien.
- . 1983 [1805/6]. *Hegel and the Human Spirit. A translation of the Jena Lectures on the Philosophy of Spirit (1805-6) with commentary*. transl. Leo Rauch. Detroit [i.o.: idem (1976 [1805/6]): Gesammelte Werke. Bd. 8: Jenaer Systementwürfe III. Hamburg].

- Honneth, Axel. 2014 [1994]. *Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte*. Frankfurt a.M.
- Knorr Cetina, Karin. 1982. "Scientific Communities or Transepistemic Arenas of Research? A Critique of Quasi-Economic Models of Science." *Social Studies of Science* 12 (1): 101-130.
- . 1999. *Epistemic Cultures. How the Sciences Make Knowledge*. Cambridge, MA.
- Latour, Bruno and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ.
- Merleau-Ponty, Maurice. 1962 [1945]. *Phenomenology of Perception*. transl. Colin Smith. New York/London [i.o.: idem: *Phénoménologie de la Perception*, Paris].
- Paßmann, Johannes. 2013. "Forschungsmedien erforschen. Zur Praxis mit der Daten-Mapping-Software Gephi." In *Vom Feld zum Labor und zurück*, ed. Raphaela Knipp, Johannes Paßmann & Nadine Taha (Hrsg.): 113-130. Siegen.
- Rieder, Bernhard & Theo Röhle. 2012. "Digital Methods, Five Challenges." In *Understanding Digital Humanities*, ed. David M. Berry, 67-84. Houndsmills.
- Schmidt am Busch, Hans Christoph (2002). *Hegels Begriff der Arbeit*. Berlin.
- Star, Susan Leigh, Geoffrey C. Bowker & Laura J. Neumann. 2003. "Transparency beyond the individual level of scale: Convergence between information artifacts and communities of practice." In *Digital library use: Social practice in design and evaluation*, ed. Ann Peterson, Nancy A. Van House & Barbara P. Buttenfield, 241-269. Cambridge, MA, London.
- Whitley, Richard D. 1972. "Black Boxism and the Sociology of Science. A Discussion of the Major Developments in the Field." *The Sociology of Sciences*, ed. Paul Halmos, 61-92. Keele.