Transparency as an Ethical Safeguard

Anna Spagnolli[1(✉)], Lily E. Frank[2], Pim Haselager[3], and David Kirsh[4]

[1] Department of General Psychology, Human Inspired Technologies Research Centre, Padua University, Padua, Italy
anna.spagnolli@unipd.it
[2] Technical University of Eindhoven, Eindhoven, The Netherlands
L.E.Frank@tue.nl
[3] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
w.haselager@donders.ru.nl
[4] University of California at San Diego, La Jolla, USA
kirsh@ucsd.edu

Abstract. Transparency seems to represent a solution to many ethic issues generated by systems that collect implicit data from users to model the user themselves based on programmed criteria. However, making such systems transparent – besides being a major technical challenge - risks raising more issues than it solves, actually reducing the user's ability to protect themselves while trying to put them in control. Are transparent systems only a chimera, which provides a seemingly useful information *pastiche* while failing to make sense upon closer examination? Scholars from ethics and cognitive science share their thoughts on how to achieve genuine transparency and the value of transparency.

Keywords: Symbiotic system · Ethics · Transparency · Implicit data

1 Transparency in Symbiotic Systems

In Human-Computer Interaction a system is considered transparent when it disappears, i.e. when it requires no specific attention from the user and leaves users free to engage in their activity. The system acquires center stage only if some problems emerge with the interpretation of its affordances and functions, an event which is considered to derive from bad design and which should be prevented as much as possible [4]. In the context of the ethics of technology, the transparency of a technology, by contrast, does not exclude the user from focusing attention on the tool, and its aim is not to facilitate performance by user or tool. Here, transparency refers to the extent to which the system discloses criteria of its functioning [3]. The goal of transparency in the context of ethics is not to enable users to effectively and easily operate a given device, it is to enable them to use such a device responsibly. The metaphor for transparency in this sense is the 'why-did-you-do-that?' button: the systems must disclose the criteria, sources, and rationales leading to its output as a way to make sure that no bias is introduced into the

J. Ham et al. (Eds.): Symbiotic 2017, LNCS 10727, pp. 1–6, 2018.
https://doi.org/10.1007/978-3-319-91593-7_1

process and to ensure that users are in a position to make informed decisions about input – giving data to the system and about output – using what the system in turn gives back.

Transparency has increasingly become the focus of many initiatives trying to deal with the ethical issues raised by systems that make decisions for us autonomously, such as those developed by Artificial Intelligence (AI), and especially those that are fed with personal information implicitly remitted to them, as is the case in what we call symbiotic systems [2]. In both these instances, transparency seems to be a critical safeguard protecting the user from unethical exploitation of their data and reducing the chance of unforeseen unethical actions resulting from system output.

There is growing appreciation that such systems must be regulated and that public bodies and professional associations must assume responsibility for such regulation. This task has been taken up by large international bodies such as, for instance, the EU in its General Data Protection Regulation, in particular Articles 12 and 13[1] or by self-regulatory guidelines from professional associations, such as the 'Ethically aligned design' guidelines by IEEE standard[2] and the Statement on Algorithmic Transparency and Accountability by ACM[3]. The idea that transparency, in its ethical sense, is one of the main safeguards of users' rights when dealing with personal data-processing technologies is dominant in all these regulations, principles and recommendations.

These efforts face several challenges. From the technical perspective, AI might work "without clear mappings to chains of inference that are easy for a human to understand" (see footnote 1); from a market perspective, a company might want to protect its own algorithms from being revealed to defend its competitiveness; from a psychological perspective, the user might be overwhelmed by information that s/he cannot manage. The very metaphor of a "why-did-you-do-that?" button is tricky. Why would a user know when a decision was or will be potentially risky? How might a user know when to press the button, or know when it is relevant to check the rationale behind the systems' output in order to make responsible decisions about it?

Identifying the challenges to genuine transparency from the user's perspective and ways to deal with these challenges was the topic of one ethical panel at Symbiotic 2017. Two scholars of ethics, Lily E. Frank (Technical University of Eindhoven) and Pim Haselager (Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen) and one scholar of cognitive science, David Kirsh (Department of Cognitive Science UCSD & Leverhulme Visiting Professor Bartlett School of Architecture, University College London) have kindly accepted to share their thoughts about this topic. Anna Spagnolli organized and chaired the panel. The following is a synthesis approved by all authors.

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[2] http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

[3] https://www.acm.org/binaries/content/assets/publicpolicy/2017_usacm_statement_algorithms.pdf.

2 What Would Ensure Genuine Transparency to Users? Panelists' Remarks

2.1 Pim Haselager: Want to Know vs. Ought to Know

I see an important aspect of transparency as being aware of what falls under one's control. I think this has two different meanings: what one wants to know and what one needs to know. The former refers to knowing what happens to my data or why the algorithm came up with a given decision. My primary example of a transparent tool in this sense would be to endow a system with some straightforward, self-explanatory system, the way a hammer's affordances are grasped. For me, the notion of affordance seems useful because it fits nicely with the non-inferential directness aimed for by symbiotic systems. The latter meaning of transparency is less discussed; it pertains to *transparency as the awareness of information required to use a system responsibly*. I make a plea that people get some sort of 'license' (not unlike a driver's license) before being allowed to use autonomous systems. You cannot simply not use certain technologies without having a reasonable understanding -on the basis of what you need to know about it- in order to deal adequately with them, make morally responsible decisions while using them and avoid e.g. negligence. What this means in practice will vary in different areas (like driving lessons vary for cars or motor cycles), for instance robotics, neurotechnology and algorithmic decision-making, and most likely also in different subareas.

2.2 Lily Frank: Transparency vs. Trust

Transparency has to do with making information available to a public or set of users regarding a particular system, but what remains vague is the quality of information to be revealed in order for a system to be considered transparent, the extent of information revealed, and level of understandability of that information. Most importantly, whether or not access to it will enable and support people to make decisions consistent with their value and priorities. I think that one the final goals of transparency should be to *expose the possible risks the user is taking and reveal the power imbalances* that already exist or are being created between the user and other institutions by use of the technology. Consequently, I see that in addition to transparency the user needs to be able trust that system. The direct link between trust and transparency should be less confidently assumed. *Trust, unlike transparency alone, includes the more general set of beliefs that the system will act in my best interests or at least consistently with my best interests or protecting my interests in a particular domain*. We should remain cautious about the burden of moral work that a concept like transparency can perform.

2.3 David Kirsh: Some Predictions About Solutions

Imagine an implant that tracks glucose levels and insulin, while also sensing and recording activity level and food consumption. To know what a person has eaten smart glasses or soon smart contact lenses would observe food choices. It's extra feature is that it can recommend items on a restaurant menu, or suggest food options from among

the possibilities in the visual field by tagging visual input. Such a device knows a tremendous about its wearer. How can someone decide whether their privacy is being breached, as in the film Being John Malkovich? How can they know what their data is worth to the corporate owners of the product? How can they know whether there are alliances or back end deals that give rise to manipulative recommendations when there is no dietary reason to prefer food A to B?

Given the closely coupled connection between device and human this implantable service might qualify as an example of extended mind. It extends the boundaries of mind to glucose and insulin monitoring, and it encodes in vision analytics of activity and person specific consequences of food choice. All in real-time. Once a person habituates to it, it becomes a part of the person: if someone else were to take it off without permission, they would commit assault and battery. It is more than just property they own.

This mind extender is one of thousands that will soon be part of our augmented selves. We will need tools to ensure that we know what is going on because of the constant threat of parasitism. This raises huge questions for law, economics and ethics.

To deal with the transparency issue one easy to imagine solution is to create *smart contracts* that provide users with fair access to those pieces of their extended mind that others are more in control of. In particular, a smart contract would allow a user, if they so wished, to follow the distribution of their personal information as it metastasizes across companies and aggregators. Who sells or shares it with whom and for what? Once information is sold by the original company what can it be used for by others? What kind of momentary payoffs, communication with third parties take place?

A smart contract is inspired by the block-chain security model, where the provenance of goods and information is kept track of as it is transferred from hand to hand. The same model could be used to track the propagation of information across services and entities of various sorts, allowing the person whose data is being shared to ask, at any moment, questions about who is receiving and using their data. This promises to partially rebalance the current information asymmetry (see Kirsh in [3]), because without such intelligible feedback only the seller of the information knows what they are collecting, how it might be used, what it is worth. The person whose information it is (or should be) knows none of these things. They are exposed and in constant danger of exploitation.

With tracking comes the possibility of a new market for bots whose job is to calculate the risk of privacy loss, the value of information, and the danger of excessive trust. Another possibility is that our own agent bots might sell our information for us. This last idea means that with smart contracting there may be new markets for buying and selling personal information and for setting smart perimeters on how far one's information can be propagated and with what risk. This is a far cry from simple transparency, for now we must trust AI agents who will act on our behalf, watching for risk in its many forms. It is our AI guardian and agent who sees the value, risk and data being collected, not us.

3 What's Next?

As Turilli and Floridi pinpoint, transparency is not yet a full-fledged ethical principle [6]; it is instead an enabling principle for successfully fulfilling some ethical principles, which might be – for instance - protecting privacy. What has been described in this panel report illustrates some strategies though which transparency may become 'enabling' to the user: the provided information must be understood by the user, limited in amount to what is necessary to make a decision and well connected to the users' goals, priorities and responsibilities. This we have considered as a genuine, *bona fide* form of transparency different from a façade transparency set up by a service provider to formally fulfill some legal obligations to be transparent. But even when it is genuine, transparency might not protect the user as much as one would expect. Acquisti et al. describe two risks deriving from making a system transparent to the user: the control paradox and the user's (unnecessary) responsibilization [1]. The former resides in users taking more risks with a transparent system because they feel they are more in control than in opaque cases. The latter consists of shifting the responsibility from the system to the user.

There is clearly much work to be done to make symbiotic systems genuinely transparent and for transparency to be ethically enabling, rather than just making some system information public. The panel goal was to pinpoint the need for work in symbiotic systems. Some additional insights partially related to transparency can be found in the papers selected for this volume from Ruijten et al., and from Gamberini et al. We otherwise refer the reader to some projects that set transparency within their main goals: the Trust and Transparency project at the Leverhulme Center for the Future of Intelligence[4]; the Global Summit on AI for Good organized by the United Nation agency specialized agency International Telecommunication Union[5]; the Partnership on IA between Amazon, Google, IBM, Microsoft, Facebook and other key players having fair, transparent and accountable Artificial Intelligence as one of its thematic pillars[6]; OPEN AI, co-chaired by Sam Altman and Elon Musk and committed to only develop safe AI solutions; and recently funded EU projects such as Types[7] and Privacy Flag[8].

References

1. Acquisti, A., Adjerid, I., Brandimarte, L.: Gone in 15 seconds: the limits of privacy transparency and control. IEEE Secur. Priv. **11**(4), 72–74 (2013)

[4] http://lcfi.ac.uk/projects/trust-and-transparency/.

[5] https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx.

[6] https://www.partnershiponai.org/thematic-pillars/.

[7] http://www.types-project.eu/.

[8] http://privacyflag.eu/.

2. Gamberini, L., Spagnolli, A.: Towards a definition of symbiotic relations between humans and machines. In: Gamberini, L., Spagnolli, A., Jacucci, G., Blankertz, B., Freeman, J. (eds.) Symbiotic 2016. LNCS, vol. 9961, pp. 1–4. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57753-1_1
3. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE (2017). http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
4. Norman, D.: The Design of Everyday Things: Revised and Expanded Edition. Basic Books, New York (2013)
5. Spagnolli, A., Conti, M., Guerra, G., Freeman, J., Kirsh, D., van Wynsberghe, A.: Adapting the system to users based on implicit data: ethical risks and possible solutions. In: Gamberini, L., Spagnolli, A., Jacucci, G., Blankertz, B., Freeman, J. (eds.) Symbiotic 2016. LNCS, vol. 9961, pp. 5–22. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57753-1_2
6. Turilli, M., Floridi, L.: The ethics of information transparency. Ethics Inf. Technol. **11**(2), 105–112 (2009)