Faith, Trust, and Pixie Dust

written by Jeremy Huggett | 02/04/2024

Adapted from original image by Kumar's Edit,
CC BY 2.0 Deed

It's been some time since I last blogged, largely because my focus has lain elsewhere in recent months writing long-form pieces for more traditional outlets. The most recent of these considers the question of trust in digital things, a topic spurred by the recent (and ongoing) scandal surrounding the Post Office Horizon computer system here in the UK which saw the false conviction for theft, fraud, and false accounting of hundreds of people. One of the things that came to the fore as a result of the scandal was the way that English law presumes the reliability of a computer system:

> In effect, the 'word' of a computational system was considered to be of a higher evidential value than the opinion of legal professionals or the testimony of witnesses. This was not merely therefore a problem with digital evidence per se, but also the response to it. (McGuire and Renaud 2023: 453)

The presumption that computers are 'reliable' was introduced into English law by the Law Commission in 1997 without any evidence to demonstrate it, and it also exists in common law jurisdictions elsewhere in the world (Mason 2021a: xv). In fact, the presumption relates to mechanical instruments but has been extended to computers without debate (Mason 2021b: 213) and legal challenges rarely succeed (Mason 2021b: 214), not least because the primary judgement of reliability frequently lies with those who program or own the system in question (Mason 2021b: 216ff), as experienced in the Horizon debacle.

In parallel, recent debates surrounding trust in computers and their outputs has been particularly associated with the growth in artificial intelligence applications, highlighted of late by the launch of the European Commission's *White Paper on Artificial Intelligence* (EC 2024) which was in part based on their earlier *Ethics Guidelines for Trustworthy AI* (EC 2019). Welcomed as placing Europe at the cutting edge of such discussions, the Ethics Guidelines identified four fundamental principles (respect for human autonomy, prevention of harm, fairness, and explicability) (EC 2019: 11-13) from which they derived seven requirements for trustworthiness: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination

and fairness; societal and environmental wellbeing; and accountability (EC 2019: 14ff). These have received robust criticism on a number of grounds, including their appearance of a 'checklist' or formulaic approach to the concept of trust and, more fundamentally, whether or not computers can be seen as trustworthy at all (e.g., see discussions in Braun et al. 2021; Freiman 2023).

Trust (and reliability as a form of trust) is not something that has been discussed in digital archaeology, and discussed more generally in archaeology largely only in passing. Trust is mostly absent from volumes on archaeological ethics (Colwell-Chanthaphonh & Ferguson 2006 being a rare example), for instance, and is at best implicit in discussions of responsibility etc., Where it is discussed at all, whether explicitly or implicitly, the focus is primarily on interpersonal trust – trust *between* archaeologists or trust *in* archaeologists, and mostly in the context of public or indigenous archaeology. Nothing about trust in our relationship with the tools that we use. And yet, trust in our digital devices is implicit in our presumptions around their reliability, consistency, accuracy, etc., but it is essentially taken for granted and hence overlooked or set aside. Is this lack of consideration healthy or wise?

There is a philosophical minefield around whether or not things other than people can be trusted, but nevertheless we frequently express our trust in devices (think satellite navigation, for instance) and are surprised when they let us down. As I've argued elsewhere (e.g., Huggett 2021), we often assign such devices agency and we regularly employ them as cognitive artefacts extending our own capabilities, and this can be linked through to trust: trust in the ability of these devices to provide us with reliable data capture and digital outputs, for instance. However, if trust in digital devices themselves is denied as no more than inappropriate anthropomorphism, for instance, then it may be that instead we place trust in the developers, programmers, and companies or organisations that created and managed the devices in the first place. I've previously suggested that ethical responsibility for the application of digital devices cannot solely be placed on the end user but there is instead an ethical chain of responsibility which connects the user with the designers, developers, manufacturers, etc., each with their own degree of ethical responsibility (Huggett 2021: 430), and trust may operate in a similar way. But whether we can have trust directly in our digital things, or indirectly in them through their makers, trust is nevertheless present in our expectations of the digital devices we use, the digital data we collect with them, and the digital repositories which serve data up to us (Huggett, in review). More on this is to eventually follow, subject to editors and reviewers!

[The title of this piece is derived from J.M Barrie: "All the world is made of faith, and trust, and pixie dust", a quote that Internet sources claim to be from Peter Pan *though I can't actually find it …]*

References

Braun, M., Bleher, H. and Hummel, P. 2021. A Leap of Faith: Is There a Formula for "Trustworthy" AI? *Hastings Center Report* 51(3): 17–22. https://doi.org/10.1002/hast.1207

Colwell-Chanthaphonh, C., & Ferguson, T. J. 2006. Trust and Archaeological Practice: Towards a Framework of Virtue Ethics. In C. Scarre & G. Scarre (Eds.), *The Ethics of Archaeology: Philosophical Perspectives on Archaeological Practice* (pp. 115–130). Cambridge University Press.

EC 2024. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (European Commission)

https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

EC 2019. Ethics Guidelines for Trustworthy AI (European Commission) https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Freiman, O. 2023. Making Sense of the Conceptual Nonsense "Trustworthy AI". *AI and Ethics* 3(4): 1351–60. https://doi.org/10.1007/s43681-022-00241-w

Huggett, J. 2021. Algorithmic Agency and Autonomy in Archaeological Practice. *Open Archaeology*, 7(1), 417–434. https://doi.org/10.1515/opar-2020-0136

Huggett, J. (in review) Trust in Digital Archaeology: Acting in an Imperfect World.

Mason, S. 2021a. Preface. In S. Mason and D. Seng (Eds.), *Electronic Evidence and Electronic Signatures* (p. xv). 5th edn. Institute of Advanced Legal Studies for the SAS Humanities Digital Library, School of Advanced Study, University of London. https://uolpress.co.uk/book/electronic-evidence-and-electronic-signatures/

Mason, S. 2021b. The presumption that computers are 'reliable', in S. Mason and D. Seng (Eds.), *Electronic Evidence and Electronic Signatures* (pp. 126-235). 5th edn. Institute of Advanced Legal Studies for the SAS Humanities Digital Library, School of Advanced Study, University of London. https://uolpress.co.uk/book/electronic-evidence-and-electronic-signatures/

McGuire, M.R. and Renaud, K. 2023. Harm, injustice & technology: Reflections on the UK's subpostmasters' case. *The Howard Journal of Crime and Justice* 62(4): 441–461. DOI: https://doi.org/10.1111/hojo.12533