HARKing to Big Data?

written by Jeremy Huggett | 18/07/2022



A 1920s aircraft detector

Big Data has been described as revolutionary new scientific paradigm, one in which data-intensive approaches supersede more traditional scientific hypothesis testing. Conventional scientific practice entails the development of a research design with one or more falsifiable theories, followed by the collection of data which allows those theories to be tested and confirmed or rejected. In a Big Data world, the relationship between theory and data is reversed: data are collected first, and hypotheses arise from the subsequent analysis of that data (e.g., Smith and Cordes 2020, 102-3). Lohr described this as "listening to the data" to find correlations that appear to be linked to real world behaviours (2015, 104). Classically this is associated with Anderson's (in)famous declaration of the "end of theory":

> With enough data, the numbers speak for themselves … Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. (Anderson 2008).

Such an approach to investigation has traditionally been seen as questionable scientific practice, since patterns will always be found in even the most random data, if there's enough data and powerful enough computers to process it.

The big data approach, where a priori hypotheses are set aside in favour of a search for serendipity, is something that has also been discussed within archaeology (e.g., Gattiglia 2015; Huggett 2020). Gattiglia essentially proposed that theory is set aside temporarily and comes back to the fore

following the data analysis, whereas I suggested theory cannot be set aside – like it or not, whether recognised or not, theory is involved at every stage from the recognition, selection, collection and recording of the data onwards (see also Huggett 2022, 276-7). Far from there being an absence of theory, theory, compounded by cultural and taphonomic processes, is implicit and influences the collection and processing of data long before hypotheses might be sought from the mining of the data. However, this still tends to assume that stepping away from the classic hypothetico-deductive method is a valid practice.

Not surprisingly, the practice has been debated long before the arrival of Big Data. Foregrounding data before theory was characterised by Kerr (1998, 197) as HARKing, or *Hypothesising After the Results are Known*. Fundamentally, HARKing is seen as problematic since it results in hypotheses that are always confirmed by the results, which is clearly a significant risk in Big Data style analyses. Subsequently HARKing has been subdivided into different types: for instance, CHARKing (*Constructing Hypotheses after the Results Are Known*), RHARKing (*Retrieving Hypotheses After the Results are Known*), and SHARKing (*Suppressing Hypotheses After the Results are Known*) (Rubin 2017, 313). None of these sound like good practice, especially the idea of deliberately (and quietly) setting hypotheses aside that don't work out. But it isn't quite so simple. Forms of HARKing are quite common in archaeological practice. We're familiar with searching for patterns within our data, deciding which ones seem to be useful, and then hypothesising about what they might mean about past lives and activities. It is, after all, related to the process of scientific induction, where conclusions are drawn from patterns observed in the data, although this is not without its problems (see Smith 2015, 19, for example, and the discussion in Ross and Ballsun-Stanton 2022).

Kerr was anxious to distinguish scientific induction from 'bad' HARKing, which he saw as presenting post-hoc hypotheses as if they were a priori hypotheses (Kerr 1998, 197). And this is where HARKing (of whatever shade) is generally seen to be problematic – when it becomes a means of disguising or mis-representing the analytical process. For example, Hollenbeck and Wright (2017, 9) distinguish between THARKing (*Transparently Hypothesising After the Results Are Known*), which they argue is acceptable, if not normal practice, and SHARKing (which they define as *Secretly Hypothesising After the Results Are Known*), which is clearly not. As they point out, many important discoveries arise through chance or error, a consequence of serendipity rather than the outcome of hypothetico-deductive reasoning:

> In fact, beyond being unable to help us uncover facts that were never recognized before, because the deductive process relies so heavily on the existing knowledge base, it may very well work *against* generating new scientific discoveries. … Thus, rejecting any findings that were not a product of a formal deductive process may limit our ability to detect new discoveries when the extant consensus in the literature is that something is impossible to anticipate. (Hollenbeck and Wright 2017, 12).

Indeed, Rubin (2022, 55) argues that the ability to judge the quality of the hypotheses, the research methods, and the statistical analyses is independent of where they are situated within the process. So there's no problem with big data mining and allied artificial intelligence related methods of analysis which hypothesise after the results, as long as we are transparent about this, then? Not entirely. If the algorithms used are black-boxed, or the internal procedures so complex as to defy understanding, and the systems incapable of explaining their reasoning (e.g., Huggett 2021,

424-428) then the transparency required of ‘good’ HARKing cannot be achieved. There is, therefore, no substitute for human intervention in the process of analysis, to understand and to evaluate the outcomes, and to determine whether the patterns and relationships identified are actually valid as well as useful. Algorithms have no way of determining whether patterns are relevant or coincidental, hence relationships between variables may be identified which are spuriously correlated with the subject of interest.

> … the more data that are mined, the more likely it is that what is found will be fortuitous, and of little or no use for understanding the past or predicting the future. (Smith 2022, 284).

Of course, human intervention is a feature of many current archaeological applications of neural networks, data mining etc., typically evidenced in a concern for ground truthing and correcting the algorithms and the patterns that the systems latch onto. But this concern is often seen in terms of training the system, implying that the responsibility for the outcomes will ultimately be transferred to the system and in the process setting aside the role of the human expert in validating the patterns and the conclusions that may be drawn from them.

As Smith (2022, 284-5) concludes:

> the ready availability of plentiful data should not be interpreted as an invitation to ransack data for patterns or to dispense with human expertise. The data deluge makes human common sense, wisdom, and expertise essential.

*[**Update:** See Brian Ballsun-Stanton’s response and thoughts about this post/subject …]*

References

Anderson, C. (2008) ‘The End of Theory: The Data Deluge Makes the Scientific Method Obsolete’, *Wired,* 23 June. Available at: https://www.wired.com/2008/06/pb-theory/

Gattiglia, G. (2015) ‘Think big about data: Archaeology and the Big Data challenge’, *Archäologische Informationen*, 38, pp. 113–124. https://doi.org/10.11588/ai.2015.1.26155.

Hollenbeck, J.R. and Wright, P.M. (2017) ‘Harking, Sharking, and Tharking: Making the Case for Post Hoc Analysis of Scientific Data’, *Journal of Management*, 43(1), pp. 5–18. https://doi.org/10.1177/0149206316679487.

Huggett, J. (2020) ‘Is Big Digital Data Different? Towards a New Archaeological Paradigm’, *Journal of Field Archaeology*, 45(sup1), pp. S8–S17. https://doi.org/10.1080/00934690.2020.1713281.

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

Huggett, J. (2022) ‘Data Legacies, Epistemic Anxieties, and Digital Imaginaries in Archaeology’,

Digital, 2(2), pp. 267–295. https://doi.org/10.3390/digital2020016.

Kerr, N.L. (1998) 'HARKing: Hypothesizing After the Results are Known', *Personality and Social Psychology Review*, 2(3), pp. 196–217. https://doi.org/10.1207/s15327957pspr0203_4.

Lohr, S. (2015) *Dataism: Inside the Big Data Revolution* (Oneworld).

Ross, S. and Ballsun-Stanton, B. (2022) 'Introducing Preregistration of Research Design to Archaeology', in Watrall, E. and Goldstein, L. (eds.) *Digital Heritage and Archaeology in Practice: Data, Ethics, and Professionalism* (University Press of Florida), pp. 15-35.

Rubin, M. (2017) 'When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress', *Review of General Psychology*, 21(4), pp. 308–320. https://doi.org/10.1037/gpr0000128.

Rubin, M. (2022) 'The Costs of HARKing', *The British Journal for the Philosophy of Science*, 73(2), pp. 535–560. https://doi.org/10.1093/bjps/axz050.

Smith, M. (2015) 'How Can Archaeologists Make Better Arguments?', *The SAA Archaeological Record*, 15(4), pp. 18–23. Available at: http://onlinedigeditions.com/publication/?m=16146&i=272889&p=20&ver=html5

Smith, G. (2022) 'The Promise and Peril of the Data Deluge for Historians', *Journal of Cognitive Historiography*, 6(1–2), pp. 277–287. https://doi.org/10.1558/jch.21156.

Smith, G. and Cordes, J. (2020) *The Phantom Pattern Problem: The Mirage of Big Data* (Oxford University Press).