Research Article

Jeremy Huggett*

Algorithmic Agency and Autonomy in Archaeological Practice

https://doi.org/10.1515/opar-2020-0136
received November 14, 2020; accepted March 13, 2021

Abstract: A key development in archaeology is the increasing agency of the digital tools brought to bear on archaeological practice. Roles and tasks that were previously thought to be uncomputable are beginning to be digitalized, and the presumption that computerization is best suited to well-defined and restricted tasks is starting to break down. Many of these digital devices seek to reduce routinized and repetitive work in the office environment and in the field. Others incorporate data-driven methods to represent, store, and manipulate information in order to undertake tasks previously thought to be incapable of being automated. Still others substitute the human component in environments which would be otherwise be inaccessible or dangerous. Whichever applies, separately or in combination, such technologies are typically seen as black-boxing practice with often little or no human intervention beyond the allocation of their inputs and subsequent incorporation of their outputs in analyses. This paper addresses the implications of this shift to algorithmic automated practices for archaeology and asks whether there are limits to algorithmic agency within archaeology. In doing so, it highlights several challenges related to the relationship between archaeologists and their digital devices.

Keywords: digital archaeology, digital practice, agency, ethics

1 Introduction

The relationship between human practice and digital tools is a rapidly developing one with dramatic changes evident across a single generation, yet the focus on tools in the present and their future development often makes it difficult to remember how much has changed since even the recent past. For example, the only personal computers available during the 1970s were limited in function and often had to be assembled from kits, but by the end of that decade they had become an increasingly mainstream consumer product. By the end of the 1980s the computer had become a truly portable and personal device in the shape of the laptop; by the end of the 1990s it could fit into the pocket (Atkinson, 2010, p. 135). The history of digital technology can be characterized in a host of different ways but in one way or another they all emphasize the rapid changes in scale, capability, and ubiquity. Relatively few digital archaeologists have witnessed this entire transformation, and "born digital" archaeologists have not experienced the before-times when computers were the size of rooms and inaccessible outside of research environments, programs ran in batch from tape, and the World-Wide Web did not exist. These different experiences add to

Article note: This article is a part of the Special Issue on Archaeological Practice on Shifting Grounds, edited by Åsa Berggren and Antonia Davidovic-Walther.

*** Corresponding author: Jeremy Huggett,** Archaeology, School of Humanities, University of Glasgow, Gregory Building, Lilybank Gardens, Glasgow G12 8QQ, United Kingdom, e-mail: Jeremy.Huggett@glasgow.ac.uk

the challenge of contextualizing the totality of the digitalization process within archaeology over time and evaluating the extent to which it has influenced and modified practice. And of course, the dramatic shifts continue with the increasing miniaturization, power, and flexibility of digital devices seen most recently in the rise of intelligent digital assistants and automated robotic devices.

Predicting the future direction of digital technologies is a thankless task since they are always in a state of developmental flux. Instead, however, it is important to consider the effects of the increasing infiltration of such devices into the practice of archaeology and the lives and livelihoods of archaeologists (Huvila & Huggett, 2018, pp. 89–90). This is not easy to do without slipping into excessively utopian or dystopian imagery, but it is necessary nonetheless, since:

> The designers of computational objects have traditionally focused on how they might extend or perfect human cognitive powers; but such objects do not simply do things *for* us, they do things *to* us as people, to our ways of seeing the world, ourselves and others (Turkle, Taggart, Kidd, & Dasté, 2006, p. 347, emphasis in original).

In particular, the development of complex algorithmic methodologies employing big data and deep learning, along with advancements in robotics incorporating sensation and dexterity have seen computers not only operate in areas formerly considered to be incapable of computerization, but to do so with little or no human intervention. In many respects, this is reminiscent of Barceló's characterization of a future automated archaeologist: a combination of mobile robotics enabling physical interaction with archaeological spaces, decision-making tools to determine best outcomes, perceptual elements linked to knowledge, and a cognitive and explanatory capability (Barceló, 2009, pp. 352–354). Similarly, Morgan's definition of a cyborg archaeology with its permeable boundaries between machines and humans which enable the digital to become "pervasive, tedious, and worryingly invisible in archaeological labor, embedded in the craft of archaeological knowledge production" (Morgan, 2019, p. 325) speaks to the growing prevalence of digital avatars, monsters, and machines as party to archaeological practice.

2 Digitalization and Archaeological Practice

In considering the place of digital devices, it is important to recognize that digitalization and archaeological practice have developed in tandem for many years. For example, a model of archaeological practice can be proposed (see Figure 1) and digital devices are implicated at each level to varying degrees. For instance, there is a long-standing tension between craft practice and standardization (e.g. Huggett, 2012, pp. 540–545) with regard to the feasibility, desirability, and extent of standardization in relation to

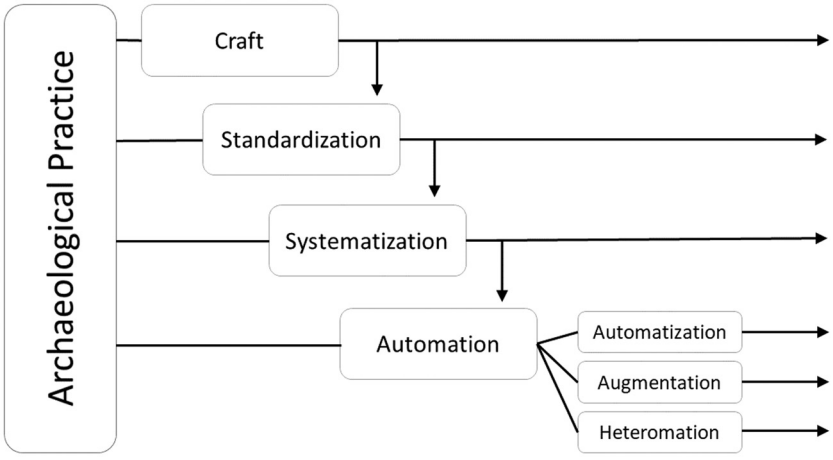

Figure 1: A model of archaeological practice.

digitalization. Most commonly, archaeological standardization operates across both data and processes: it may entail data content standards, data documentation standards, and ontology standards, together with the definition of tasks and procedures to enable them to be undertaken in a reproducible manner, and so may consequently be seen to substitute creative craft with formalized methods (e.g. Caraher, 2016). This routinization of practice can typically lead to its systematization through the development and use of digital tools and technologies which support a broad range of simple through complex archaeological tasks (ranging from spreadsheets to geographical information systems, for instance) which frequently rely on varying degrees of standardization. In turn, the systematization of practice can facilitate automation through the delegation of functions and tasks to digital devices, although automation may be elaborated further (see Figure 1). Automation may incorporate augmentation, where digital devices provide assistance rather than outright replacement, incorporating digital cognitive artifacts (e.g. Huggett, 2017) and robotic devices (increasingly used in survey and laboratory practice, for example). A second variant is automatization, where combining information and augmentation enables the development of data-driven modelling, classification, and artificial intelligence applications (e.g. Dertouzos, 1997, pp. 83–84). A third category of automation, heteromation (Ekbia & Nardi, 2017), flips the human-digital relationship with people performing free or low-cost tasks to support the technological system, evident in the development of microwork systems and crowd-sourcing applications in archaeology (e.g. Bonacchi et al., 2014; Ridge, 2013).

There is no straightforward evolution in this model of practice; instead, there is considerable overlap between the different categories such that any may function at one or more stages depending on the resources of the practitioner. From a purely practice perspective, therefore, archaeological practices that are otherwise contemporary may operate to greater or lesser degrees in different categories. For example, excavation recording may be seen to operate at the craft level, perhaps manifested in the field notebook or site book. Equally, it might function as a standardized practice (evidenced by pro-forma recording sheets) or systematized practice (evidenced by the use of digital tablets etc.). From a practice perspective, standardized practice does not need to necessarily incorporate craft practice (as seen in debates between craft practitioners and proponents of standardization (e.g. contributions to Cooper & Richards, 1985)), and indeed standardization is frequently seen to introduce efficiencies in practice through reducing the creative aspects of craft practice (e.g. Caraher, 2016).

On the other hand, there is a more evolutionary aspect when traversing this model from a digital perspective: systematization is predicated on standardization, while automation requires both standardization and systematization. For example, creating a database for excavation recording requires standardization of categories (fields) and ideally a standardized set of terms. Systematization requires these to be in place for a successful implementation of a recording system, while automation of any kind requires all the preceding elements to be in place, establishing a chain of dependency.

A third way of traversing the model is to consider the changing relationship between practitioner and digital: for example, the human archaeologist may be seen as dominant in craft practice whereas in automated practice the digital device assumes a greater degree of authority and control. This in turn raises the question of where the balance of agency is located in digital archaeological practice: does it remain with the archaeologist throughout, or does the power to make significant changes shift towards the digital at some point?

3 Agency and the Digital

Whether or not things – digital or otherwise – can have agency is something that is widely debated, both in archaeology and elsewhere. As Dobres (2020, p. 76) notes, the concept of agency in archaeology has been defined and applied in a number of competing ways and further complicated by differing concepts of "things." At its simplest, agency is broadly defined as the capacity to act, emphasizing the importance of intentional meaningful action (e.g. Gell, 1998, p. 16; Robb, 2010, p. 513 and 515; Schlosser, 2019). Such definitions tend to implicitly associate agency with human action, reliant on consciousness and will, and

consequently non-humans and especially inanimate objects are without intent and hence have no agency. Hence Broussard (2018, p. 89), for example, argued that even a machine that can "learn" and thereby improve at a programmed, automated task, does not acquire knowledge, wisdom or agency. Such a device has no intent, responsibility, or liability (an issue in relation to the legal status of self-driving cars, for instance), and so is not (yet) truly independent. In archaeology, Ribeiro, for example, has similarly argued that agency entails freedom to choose to act and those actions constitute responsibilities requiring ethics, so consequently agency cannot be assigned to inanimate objects (Ribeiro, 2016, p. 231). In a more nuanced discussion, Lindstrøm maintained a separation between human and object on the basis of intentionality, arguing that inanimate things might have at best a "secondary," "reactive," or "distributed" agency derived from their human associations, or else agency might simply be applied metaphorically to objects (Lindstrøm, 2015, pp. 227–228). Indeed, this kind of approach to agency had for some time been seen in archaeology as a means of accessing the intentionality of the people behind the physical archaeological evidence through the "secondary" agency imbued within their artefacts (following Gell (1998, pp. 20–21), for example). This separation into agency as human intentionality and agency as things acting upon (and with) people was summarized by Robb (2015, p. 168) as people possessing the agency of "why" whereas things have the agency of "how."

Such approaches to agency largely retained the focus on the human subject such that non-humans and objects owed their agency to human intention and action, and resistance to this can be found in various postmodern and posthumanist approaches. For example, Haraway's classic concept of the cyborg (e.g. Haraway, 1991, p. 149ff) blurred the boundaries between human and things, while for Barad, agency was not a peculiarly human characteristic but something that emerged as a consequence of interaction: "… if agency is understood as an enactment and not something someone has, then it seems not only appropriate but important to consider agency as distributed over nonhuman as well as human forms." (Barad, 2007, p. 214). In archaeology, symmetric approaches (e.g. Olsen & Witmore, 2015; Witmore, 2007) and actor-network theory (e.g. Knappett & Malafouris, 2008; Olsen, 2010) similarly sought to decenter the human and adopt a less anthropocentric approach to agency through emphasizing the agency of inanimate things and non-humans. This posthumanist approach sees agency distributed beyond the human with the human becoming a participant rather than necessarily the instigator of action, although privileging the agential status of things can be seen as creating a correspondingly dehumanized archaeology (e.g. Damilati & Vavouranakis, 2021, p. 122) in which humans are marginalized and objectified. Hodder's view of agency challenged a perception of equality or symmetry between humans and things, (e.g. Hodder, 2015), instead seeing agency in terms of entanglement where humans depend on things, things depend on other things, things depend on humans, and humans depend on humans (Hodder, 2012, p. 88ff), all at the same time. Hodder's entanglement placed a key emphasis on dependence and dependency: the way that humans and things enable as well as constrain each other's actions in a combination of both symmetrical and asymmetrical relationships which entrap both (Hodder, 2012, pp. 92–94). Barrett also rejected the symmetry between humans and things, arguing that, unlike machines, humans work towards their growth and renewal through participating in a changing assemblage of both things and other people (Barrett, 2014, p. 71). However, digital things may present a greater challenge to a humancentric view of agency, particularly when some are themselves capable of self-improvement if not renewal, foster social and political change (e.g. Kaufmann & Jeandesboz, 2017, p. 310), and able to empower as well as disenfranchise long before any suggestion of a technological singularity is reached (e.g. Vinge, 1993).

Not surprisingly, technological and archaeological approaches to agency broadly mirror each other given their often-shared theoretical foundations. However, it may be that the development and use of digital objects changes the terms of these traditional debates in certain key respects. Hayles (2005, p. 177), for example, observed that the classic image of agency was muddled by perceptions of the human as a biological machine (in relation to understanding the functioning of the brain, for instance) or a machine as similar to a biological organism (as in computer reasoning as a model of the human mind). If, for example, humans were like machines then agency could not be securely located in the conscious mind, whereas if machines were like biological organisms then they must possess agency even though they

are not conscious (Hayles, 2005, p. 177). The consideration of cognition can therefore extend agency beyond humans to non-human life forms and to advanced technical systems (e.g. Hayles, 2017, pp. 30–31), both of which may possess specialized cognitive capabilities that are superior to humans. While this might seem to deny agency to things other than the most advanced forms of artificial intelligence, it broadens the scope of agency by allowing consideration of the concept of cognitive artifacts. Cognitive artifacts are not necessarily capable of cognition themselves: they support human cognition through providing a mixture of techniques, tools, calculations, and interventions which enable humans to offload laborious or complex tasks onto external devices. In archaeology, for example, digital cognitive artifacts may include digital cameras, total stations, laser scanners, as well as computers (Huggett, 2017, Section 2). Cognitive artifacts are associated with the extended mind thesis (e.g. Clark & Chalmers, 1998): the idea that a proportion of a cognitive task might be conducted beyond the human agent by an external device. Whether or not such a device is itself capable of cognition is contested; however, cognitive artifacts may be more simply seen as scaffolding or complementing human cognition, providing facilities that are beyond normal human capabilities but without possessing cognition themselves (e.g. Menary, 2010; Sutton, 2010). This distinction between two forms of cognitive technology is also found elsewhere: for example, Reiner and Nagel distinguish between technologies of the extended mind in which algorithmic output is seamlessly integrated with the human mind (illustrated by ultimate belief and trust in a satnav device), and cognitive support, where the algorithm and its results are not integrated with the human mind but instead scaffold human cognition (Reiner & Nagel, 2017, pp. 110–111). In archaeology, for example, digital cognitive artifacts support archaeological cognition through providing the capability to see beneath the ground without disturbing it, or to characterize the chemical composition of materials, for instance (Huggett, 2017, Section 3), both of which are beyond human capacity, and do so without presuming that the devices themselves necessarily possess cognition.

If it can be argued that agency may be associated with cognitive artifacts which perform their own cognition, following the extended mind thesis, the question remains whether cognitive artifacts which simply complement human cognition without having cognition themselves can also possess agency. To archaeological proponents of object agency, this might seem to be self-evident. If, for example, the wielding of Sørensen's hammer (2016, p. 119) or Malafouris's walking stick (2015, p. 357ff) convey some degree of agency, then it would be unreasonable to think that a digital device (whether an instrument or a piece of software) does not. Moreover, digital agency may be broken down further. For example, Krakauer (2016) distinguishes between complementary cognitive artifacts which support human cognition via the creation of mental models, and competitive cognitive artifacts which essentially replace human cognition such that, taken to its ultimate conclusion, the artifact reduces the human capacity to operate in those areas. For instance, he distinguishes between the support provided by an abacus or slide rule, suggesting that after repeated use they could be set aside and the mental model used effectively in its place, whereas the mechanical/digital calculator or GPS device acts as a replacement through amplifying human capacity while frequently reducing human capability to perform the task in the absence of the device (Krakauer, 2016). This categorization may be refined further: for example, Rammert (2012) defines different levels of agency on the basis of their technical mode of operation and degree of autonomy (see Table 1), with rising

Table 1: Levels of technical agency. Adapted from Rammert (2012, p. 97)

Level	Mode of operation	Examples
1	Passive	Devices entirely operated externally, such as digital calipers or laser levels
2	Semi-active	Devices having some form of self-operation, such as a total station or magnetometer
3	Re-active	Adaptive devices with feedback mechanisms, such as 3D laser scanners
4	Pro-active	Devices with self-activating programs responding to data, such as machine learning systems
5	Cooperative	Devices with high degree of independence and self-coordination, such as autonomous vehicles for sonar and optical survey, and bio-mimetic robots

degrees of mobility, context sensitivity, programmability, and communication enabling them to be observed as increasingly pro-active rather than passive actants (Rammert, 2012, p. 96).

Ultimately, whether or not agency in the sense of a capacity for intentional and/or cognitive action can be legitimately associated with digital devices, agency can certainly be attributed to devices by humans, especially given the tendency to anthropomorphize them. For example, Wegner argues that when humans project action to imaginary agents, they create virtual agents which appear to be capable of action, and treating such imaginary agents as real reinforces a sense of engagement (Wegner, 2002, p. 221 and 228ff). On that basis, if a device affects subsequent human actions and decisions then it can be said to have agency, even if that agency masks the human agency involved in the design and creation of that device. For example, agency may be inscribed into the algorithms governing the functioning of digital hardware and software devices by human programmers which may lead to algorithmic bias (e.g. O'Neil, 2016), even if that algorithmic bias is primarily a reflection of human bias (Hill, 2018, p. 12). Furthermore, human actors can be seen to share agency with devices where the task could not be done without the participation of the non-human components (e.g. Hanson, 2014, p. 60) in what may be characterized as a symmetric or asymmetric relationship (Huggett, 2017, Section 3). Evidence of machine agency can also be sought in the responses of the human practitioners themselves: adapting practice to the new technology, adopting workarounds to perceived shortcomings in the digital tools, and responding to increased demands in terms of the speed, volume, and complexity of the task as a consequence of the introduction of a new technology, for instance (Huvila & Huggett, 2018, pp. 91–94; Woods & Dekker, 2000, pp. 273–274).

In short, it can be argued that to suggest that a digital device has no agency at all would be a form of category error: it is safer to work on the assumption that it does possess some degree of agency and work through the implications associated with that than to ignore the possibility and be blind to the potential consequences. So if a device affects subsequent human actions and decisions then it can be said to have agency, or if a device performs a task that could not easily be done without the non-human contribution, then it can be argued to at least share agency with the human actor. In summary, if there is an approximate equivalency between a human and a digital device in the performance of a task, then we can claim agency.

4 Digital Agency in Practice

One of the key effects of such cognitive agential devices is not simply to reproduce practice – at least in theory making it faster, more efficient, more reliable, more replicable – but to change it in the process. New technologies are frequently claimed to be disruptive (online streaming services leading to the demise of high street video rental stores and the decline of film DVDs and music CDs, for example), making change a seemingly inevitable outcome of technological innovation. Practice is expected to adjust accordingly, but this is not always the case: there may be unintended side effects, unexpected complexities in application and use, and shortcomings or failure resulting from poor adaptation or design (e.g. Huggett, 2000; Woods & Dekker, 2000, p. 274). Human responses to these practical deficiencies can vary from outright resistance to developing workarounds which hack the designed systems, subverting their scripts, processes, and procedures to make them usable in acts of "covert agency" (Applin & Fischer, 2015, p. 1). For example, early attempts to introduce on-site computer recording to an archaeological excavation met with resistance from site supervisors who found that the digital tools inadequately compensated for the loss of traditional paper-based recording, and the presence of computers was seen to reduce efficiency through information and communication overload (McVicar & Stoddart, 1986). Similarly, the tedium created by the computer recording system developed by the All American Pipeline Project led to archaeologists developing unspecified but sophisticated means of sabotaging the system (Plog & Carlson, 1989, p. 263). An experiment to use digital pens on the Silchester Town Life Project to speed up the transfer of site context records into the site database was dropped as supervisors found them difficult to integrate into established working practices and there was little benefit found from their application (Rains, 2015, pp. 82–84). A discussion of tablet computers used for on-site recording noted how archaeologists could become detached from the physicality

of the site through their focus on the screen, and there was a tendency to defer primary recording to an off-site activity when faced with a range of practical and technical issues (Taylor et al., 2018, Section 10.1). In such ways, digital agency is revealed, albeit not necessarily in a positive light.

In part, the range of hacks, subversions, acts of resistance, and other forms of covert agency reflect conflicting expectations over whether current methods can be simply embedded within digital devices, or whether those practical methods require re-evaluation and reshaping in order to capitalize on the introduction of the technological tools. For example, Structure from Motion imagery offers the potential to radically improve the speed, accuracy, and flexibility of archaeological on-site data recording and hence might replace traditional methods in pursuit of commercial efficiencies or else be a means of allowing the archaeologist more time to focus on interpretation drawings and to maintain a closer engagement with the physical evidence (Powlesland, 2016, p. 28).

This underlines the tension that exists between human (archaeological) agency and digital agency and the balance that may or may not be struck in different situations. Underlying this unease is the sense that digital technologies may subvert and subdue human decisions, even to the extent that humans may be shaped and used by the technology. For example, "... instead of the IT artifact being shaped and used by humans, humans can actually be considered as "artifacts" being shaped and used by machines" (Demetis & Lee, 2018, p. 930), with humans reacting to technological stimuli rather than technology reacting to human stimuli. Demetis and Lee (2018, p. 944) illustrate this through a visual metaphor, graphing human against technological agency (see Figure 2). Where there is high human agency and low technological decision-making, there are strong human/technology interactions and weak technology/technology interactions. As human agency reduces and technological agency increases, a theoretical point is reached where the roles reverse and technological agency becomes more significant, with weaker human/technology interactions and stronger technology/technology interactions. They also include a futuristic scenario where both high human and technological agency are combined in a potentially symbiotic human augmented with artificial intelligence (Demetis & Lee, 2018, p. 944). In terms of archaeological practice, most digital practice still incorporates relatively high levels of human agency and hence sits in the upper left area of the graph (see Figure 2). However, the development of automated tools for classification and identification is shifting the balance towards the lower right area, with increasing levels of digital agency. Whether or not

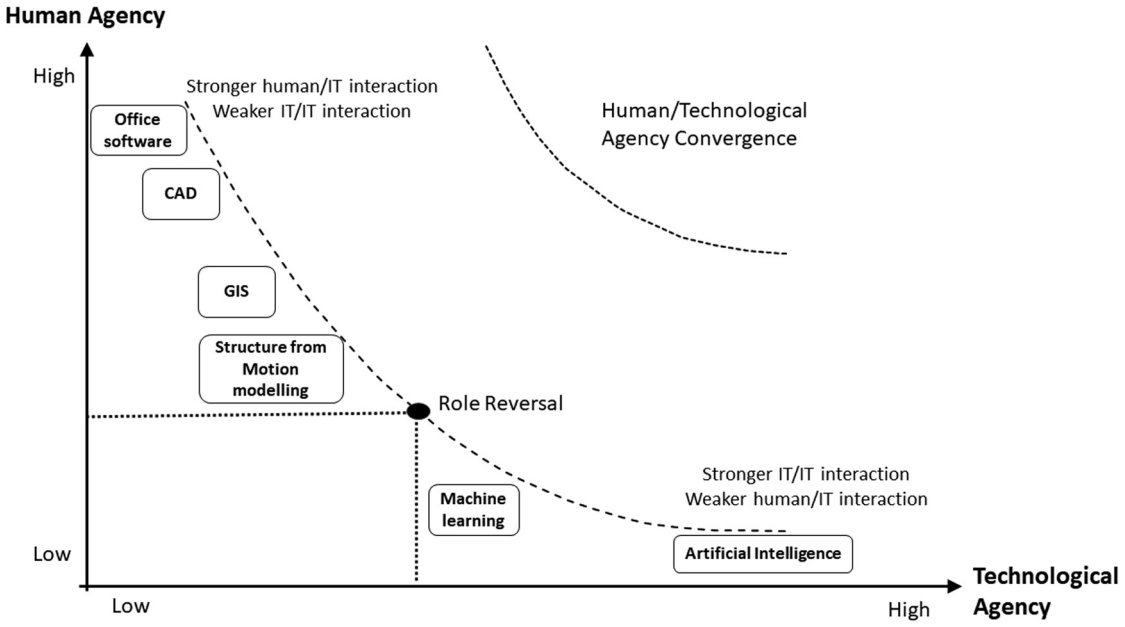

Figure 2: The relationship between human and digital agency in archaeological practice. Adapted from Demetis and Lee (2018, p. 944). Examples of digital applications in archaeology illustrate the shifting balance of human and technological agency.

archaeological practice has reached the tipping point in certain areas where we can identify role reversal remains debatable, although a number of artificial intelligence/machine learning applications explicitly claim to offer the prospect of replacing, rather than supporting, human experts, which would move practice beyond this tipping point.

This model has significant implications for our approach to implementing digital devices and employing them within archaeological contexts, as well as presenting a number of ethical issues. Altering the balance of agency constitutes shifts in power and authority, determining where the core of archaeological practice in terms of recognition, decision-making, action, recording, analysis, and interpretation sits. As Figure 2 highlights, where the balance of agency sits is also linked to the exercise of autonomy, and there is a growing body of increasingly autonomous digital systems contributing to archaeological practice, ranging from archival search engines through aerial and underwater drones, laboratory-based robotic devices, to machine learning systems for classification and feature detection, for example. Hayles defines this growing autonomy as one of "punctuated agency," consisting of long periods where human agency is critical, and shorter intervals when the systems operate without human intervention (Hayles, 2017, p. 32). For example, an aerial photographic drone has punctuated agency in the sense that it is set up by the archaeologist and its flight path defined, but once it is airborne it manages its own flight, including collision avoidance, captures the images, and returns to its take-off location as batteries fail or the task is completed. Thereafter, agency returns to the archaeologist who downloads the images and loads them into software for photogrammetric processing, at which point another set of algorithms take over the task. This image of punctuated agency appears to carry reassurance that human agency retains authority without surrendering control, with devices used primarily to make sense of information, to perform calculations, and to issue results. Human control can seem to be retained through choosing to use the devices in the first place, selecting the procedures, as well as accepting or rejecting their outputs, although the extent to which the devices imperceptibly manipulate or coerce their users (e.g. Danaher, 2018, p. 640), and reduce human autonomy through largely invisible influence, remains open to question (e.g. Onsrud & Campbell, 2020, p. 237). Although this concern is commonly linked with artificial intelligence and other allied techniques, it would be mistaken to assume that the effects of automation do not reach back into areas which are less obviously associated with digital cognition and reasoning and where consequently the implications of automated actions remain largely hidden.

5 Black Box or Glass Box?

The metaphor of the black box has become almost a cliché in discussions concerning automation. Classically, black boxes are devices whose inputs and outputs we can observe, but the means by which those inputs become the outputs is not required or not able to be determined (Latour, 1999, p. 304). Fundamentally, black boxes generate actions – and hence have agency – but knowledge of how they arrive at their outcomes remains hidden. They are seen as impenetrable, inscrutable, powerful entities that "rule, sort, govern, shape, or otherwise control" (Ziewitz, 2016, p. 3), part of a "data-driven algorithmic culture" (Striphas, 2015, p. 396) or even a "computational theocracy" (Bogost, 2015). Users become accustomed to manipulating systems whose core assumptions are not understood, and seduced into abdicating their authority (Turkle, 1997, pp. 36–42). Black-boxed procedures and practices make the intervening processes opaque, with the degree of invisibility often being perceived as a measure of success. It is often only when a device stops working or behaves unexpectedly that it reveals itself as constituting one or more black boxes. Many of the devices now habitually used as part of archaeological practice frequently behave as black boxes: we know the inputs into our digital cameras, total stations, laser scanners, proton magnetometers, X-ray fluorescence machines, and so on, and we use their outputs in our records and our interpretations, but we often know little about their internal workings and arguably do not need to in order to make use of them until they become unreliable, misleading, or fail in some way (Huggett, 2017, Section 7).

Much of the discussion of black-boxing implies that this is something particularly associated with digital devices, whereas it is equally a feature of human practice. For example, tacit or habitual practices

are frequently characterized by an absence of explanation or discussion, whether deliberately or inadvertently. Black-boxing may simply arise through familiarity: for instance, Leighton (2015, pp. 69–69) points to radiocarbon dating and excavation methods as examples of closed debates taken largely on trust. Methodologies or techniques are often not articulated because they are commonplace, customary, or assumed to be self-evident; furthermore, the decisions behind actions may be effaced and consequently underlying limitations may be misunderstood or meaning lost (e.g. Huggett, 2020a, pp. 7–8). The difficulty that human experts can have in clarifying what may otherwise be black-boxed inferences and tacit knowledge is a reminder that challenges in unboxing practice are not only associated with digital devices. Recognition of this can be used as an argument for not opening the digital black box in the first place and evaluating performance primarily on the basis of outputs (e.g. Jones, 2018), since to do otherwise would set the explanatory bar higher for black-boxed tools than for human experts. An alternative approach is to see explanations or justifications of outcomes as only necessary when those outcomes are not crucial or where there are no unacceptable consequences (e.g. Doshi-Velez & Kim, 2017, p. 3). Generally, however, it is accepted that for trust to exist a digital black box must be capable of being translated into a glass box through which what was previously hidden is revealed (e.g. Gilpin et al., 2018; Guidotti et al., 2018). This then shifts the problem onto the question of just how much transparency is enough. In a sense, this parallels the question of reproducibility in Open Science (e.g. Marwick, 2017) in terms of what is sufficient for an appropriate level of interpretability. For example, is the provision of source code feasible given the commercial nature of many of the devices used in archaeology, and if so, how much (the code in its entirety or just the key algorithms, raw code versus pseudocode, etc.)? Digital devices are embedded with diverse theories and practices but do these need to be available to a user to establish agreement prior to application and use? Where a digital device appears capable of explaining its reasoning, such as an explainable artificial intelligence providing some visibility of its underlying process, is it in reality creating a new black box through supplying a gloss that is understandable by the user? In many respects, these tensions create an "anthropocentric predicament" (Humphreys, 2011, pp. 134–135) where computational methods and devices exceed human abilities and operate in ways that humans cannot fully understand, so how can they be appropriately evaluated? Consequently, the search for transparency becomes an ethical problem when verification is not possible as a result of reliance on devices whose functioning is not fully understood (e.g. Dennis, 2020, pp. 212–213).

Many of the reasons why digital devices should be glass-boxed rather than black-boxed are paradoxically also the reasons why this is a difficult thing to do. For example, abstraction underpins all digital applications: in one way or another they take a complex system from the real world and abstract it into processes that capture some of that system's logic and discard others. These underlying algorithms are linked with other processes to create software, which are themselves abstractions. The effect of this abstraction at multiple levels – algorithm, code, software – is to distance the user (the designer, the programmer, the end user) from the need to be concerned with the low-level binary operation of the computer. As a result, at the same time as they serve our purpose these tools constrain and limit in ways that may not be fully appreciated because of the opacity they impose. For instance, there is the opacity associated with proprietary software (Burrell, 2016, pp. 1–2), concealed by patent and copyright to ensure corporate protection. It restricts what can be done with the software, denying disassembly and even ownership. The response to this is to make the code openly available for scrutiny and audit but in turn this encounters opacity in the sense that reading and writing code is a specialist skill, and one outcome of increased abstraction is that fewer need to know how to code: a form of opacity through technical illiteracy (Burrell, 2016, p. 4). Even given the access, ability, and time to be able to understand the complexities of the underlying code, it may remain impenetrable as a consequence of the opacity of the operation of the device (Burrell, 2016, p. 5) and the size of the codebase (e.g. Christin, 2020, p. 3). Machine learning techniques are especially problematic in this regard as they proceed by constructing representations without reference to or regard for human understanding, and consequently "machine optimizations based on training data do not naturally accord with human semantic explanations" (Burrell, 2016, p. 10). This would suggest that the argument for the benefits of using machine learning in archaeology based on the need to quantify and explicate the different variables and threshold values and making the process replicable (Davis, 2020, p. 3) is not the whole story.

6 Accessing Black Boxes

A range of ways of accessing these agential black boxes have been suggested although none are without their challenges. For example, Huggett (2017, Section 8) proposed a layered series of approaches to unpacking digital devices, starting with a critical reading of the code, then a critical appreciation of the software system itself, then a critical engagement with the creative process its design, implementation and development, and finally a critical understanding of the use of the hardware/software within its archaeological context. Similarly, Christin (2020, p. 3ff) identified three approaches: algorithmic audits, cultural and historical critiques, and ethnographic studies. In particular, she proposed an ethnographic toolkit of practical strategies to bypass algorithmic opacity, employing refraction (examining the changes that take place when algorithms are used) (Christin, 2020, pp. 10–11), comparison (examining similarities and differences between algorithmic systems) (Christin, 2020, pp. 11–12), and triangulation (addressing data sampling, reflexivity, and social connections (Christin, 2020, pp. 12–14). Fundamentally, Christin argues for the importance of working with algorithms to bypass their opacity (Christin, 2020, p. 16) and in this regard echoes arguments for similar approaches to understanding digital devices and their operation in archaeology (e.g. Huggett, 2012, pp. 546–548; Huggett, 2017, Section 8), ethnography (e.g. Seaver, 2017, 2018) and in anthropology more widely (e.g. Horst & Miller, 2012). All share a common focus on indirectly looking at the digital device by examining its inputs, outputs, contexts of design and use, effects of use, etc. without opening the black box itself. As Bucher suggests, when confronting a black box the importance lies in examining the properties that can be discovered and identifying those which may be obscured, rather than necessarily knowing the detail of what is inside (Bucher, 2016, p. 86). An indirect approach therefore directs attention to the messiness that the black box seeks to hide (Bucher, 2016, p. 94).

This indirect focus has much to recommend it, not least in terms of its feasibility compared with attempts to unbox the black box of digital devices. But is it sufficient? Is it adequate to know the inputs and the outputs without information about the flows and manipulations in-between? Is an understanding of the intentions of the designers and developers enough to enable a confident application of the tools they have designed, especially when, as is common in archaeological practice, the tools themselves have been taken out of their original context and applied in areas that were not part of the original design parameters? Is it satisfactory to accept that the black box, even if it were capable of being opened, is beyond comprehension because of the number of variables and the complexities of the algorithms and their interrelationships? Can it truly be claimed that methods and results of analysis are open and transparent if the devices used remain opaque?

In some respects it would be simple to argue that not all digital devices and their algorithms are equally required to be transparent: for instance, tools that have become embedded and customary in practice have perhaps passed the point at which this is either useful or necessary. We do not need to know the low-level details of how a computer works, or how a word-processor helps us compose our thoughts, to be able to make successful use of them. Nevertheless, lack of knowledge about the functioning of agential digital devices raises ethical and practical questions, be they digital survey instruments (e.g. Huggett, 2017, Section 5) or digital cameras (e.g. Dennis, 2020, pp. 212–213), for instance, if only through the risk of uncritical or naïve use. Realistically, however, the fact that most of these devices are commercial products makes glass-boxing their internal procedures an unlikely prospect, and an indirect contextual approach to such black boxes seems the only feasible option.

This may be a reasonable and inevitable approach to non-cognitive devices on the left-hand side of the "reversal" point in Figure 2, where human agency is uppermost, but the shift to greater degrees of technological agency and competitive cognitive artifacts gives rise to concerns about digital practices and autonomy lying beyond human control, and hence an approach which simply focuses on inputs and outputs and contexts of use is neither reasonable nor should it be inevitable. That said, it is also the case that such devices have not yet become part of day-to-day archaeological practice and are instead at present primarily found in research contexts. Consequently, this offers an opportunity to address the challenges posed by autonomous digital tools possessing technological agency before they are widely implemented within the discipline.

7 Explanation, Interpretability, and Comprehensibility

If agency is to be shared between human and digital device and/or technological agency is to exceed human agency, comprehensibility of the black box interior becomes a crucial question. For example, the creation of Explainable Artificial Intelligence (XAI) recognizes the importance of providing explanations and details of its functioning to be suitably intelligible to its audience (e.g. Barredo Arrieta et al., 2020, p. 85ff). Access to an explanation is seen to provide accountability and empowerment, actionable information enabling decisions about whether different outcomes might be achievable, and a basis for evaluating the validity and justifiability of the outcomes (Selbst & Barocas, 2018, p. 1118ff). This therefore goes considerably beyond knowing inputs and outputs and context: it requires all the internal features and processes of the black box system to be known and understandable, and thereby presents a considerable challenge.

The search for explanation combined with interpretability in digital black boxes is not unsurprisingly sought through technological solutions (e.g. Barredo Arrieta et al., 2020; Gleicher, 2016; Guidotti et al., 2018; Selbst & Barocas, 2018), although these are far from straightforward. For example, early forms of expert system used knowledge bases consisting of a collection of rules which meant that the set of rules triggered, and the consequent chain of inference, was able to provide an explanation for the conclusions drawn. In contrast, the complexity of deep learning models and neural networks with potentially hundreds of layers and thousands of parameters, creating relationships that are not pre-defined and intermediate data that are not interpretable, make such systems highly opaque even to their designers. For example, most archaeological applications of machine learning employ pre-trained networks such as ResNet for image classification. The ResNet network is based upon the generic ImageNet dataset which consists of over 14 million annotated images in more than 20,000 categories and sub-categories (Yang, Qinami, Fei-Fei, Deng, & Russakovsky, 2020, p. 549). ResNet incorporates approximately 5×10^7 learned parameters and carries out around 10^{10} calculations to classify a single image, and each layer within the ResNet network computes between 64 and 2,048 channels of information per pixel (Gilpin et al., 2018, pp. 82–83). Among the archaeological applications of ResNet, the ArchAIDE neural network for pottery classification used ResNet-101 with 101 layers (Itkin, Wolf, & Dershowitz, 2019, p. 9), while the WODAN system for feature recognition from airborne laser scanning data used ResNet-50 with 50 layers (Verschoof-van der Vaart & Lambers, 2019, p. 36), and the neural network used to detect cultural heritage features on Arran used ResNet-18 (Trier, Cowley, & Waldeland, 2019, p. 169). With even the smallest of these, such figures underline the complexities of overcoming opacity, especially when it is considered that these systems are currently for the most part only capable of identifying or classifying a handful of categories of artifact or site. The fact that such pre-trained neural networks use non-archaeological images as the basis for image classification with archaeological data only being added in the final stages of model building provides an added complication, as is explicitly noted in some cases (e.g. Trier et al., 2019, p. 173).

Selbst and Barocas (2018, p. 1110) identify three main approaches to explainability in machine learning: building models that have explainability build into them from the outset, post-hoc methods which approximate the model in a way that is more easily explainable, and interactive methods which enable a clearer understanding of the functioning of the model (see also Barredo Arrieta et al., 2020; Gleicher, 2016, for example). Each set of methods incorporate trade-offs between comprehensibility and accuracy, interpretability and completeness (Gilpin et al., 2018, p. 81), and entail considerable simplification. For example, designing explainability into a model from the start may require only a limited set of all possible variables to be considered, or the use of a more transparent method such as decision trees which may need an upper limit on the number of branches and leaves in order to remain comprehensible (Selbst & Barocas, 2018, pp. 1110–1112). Post-hoc explanations may involve textual explanations or visualizations which provide simplifications or summaries of sections of the model, or simplified models of the model which use methods such as rule extraction or decision trees (e.g. Barredo Arrieta et al., 2020, p. 88). Interactive methods enable the user to iteratively change parameters to see their effect on the outcomes, or identify features which have the greatest effect on the results (Selbst & Barocas, 2018, pp. 1114–1115), while leaving the algorithms themselves or the underlying model as a black box (Gleicher, 2016, p. 83). Indeed, the search for solutions may result in the explanations themselves being black-boxed, making it difficult to recognize when

explanations that are plausible are actually misleading or biased, and in that respect no different to human explanations (Lipton, 2018, p. 43). In fact, the opacity and problems with comprehension inherent in these devices serve to highlight existing equivalent problems with human expertise (Amoore, 2019, p. 150): the outputs of complex systems such as these are just as culturally, socially, politically, and technologically situated and contingent as both the data used in their creation and use, and the humans developing and using the systems.

The limitations of these approaches do not deny the purpose or importance of explanation. An understanding of the model is crucial for its developers, since if an output is unexpected or in error the reasons for this need to be found and the problem corrected. Explanations as a means of justification and accountability are if anything even more significant if the system is to become embedded in practice, where the acceptance and confidence of end users who are typically unconnected with the development of the device will, at least initially, be reliant upon the transparency of operation and justification of the conclusions. At present, research within archaeology is primarily focused on the development of these tools with less effort currently expended on incorporating methodologies for explanation of their results, but in future this situation is likely to change, accompanied by a corresponding demand for appropriate levels of explanation.

8 Balancing Agency and Autonomy

Although many of these systems are primarily research tools rather than intentionally developed for incorporation within archaeological practice, the precise relationship between human and technological agency and autonomy remains unclear in most cases. For instance, most applications of machine learning in archaeology currently focus on the identification and automated classification of artifacts (primarily pottery but also lithics), or on the automated identification and classification of features from aerial or satellite imagery. Their proponents principally see these tools as complementing and supporting current archaeological practice (e.g. Bennett, Cowley, & De Laet, 2014, p. 897; Makridis & Daras, 2012, p. 3; Trier et al., 2019, p. 168; Wright & Gattiglia, 2018, p. 61) which suggests that agency may be shared but ultimately human control is retained. However, in some cases the system can appear to replace human expertise. For example, Arch-I-Scan, a classification system for samian ware, is described as capable of use by people with relatively low levels of expertise so that experts can focus on more analytical and less mundane tasks (Tyukin et al., 2018, Section 4). Even if such tools are initially targeted at augmentation, they may inadvertently lead to the substitution of human expertise and the consequent transfer of agency. It is a relatively small step from a system developed to speed up the coverage of large area mapping of heritage sites and thereby support the archaeological surveyor (e.g. Trier et al., 2019, p. 168), or a system designed to support archaeological specialists through the automated classification and interpretation of pottery sherds (e.g. Wright & Gattiglia, 2018, p. 61), to a system which replaces human expertise in the pursuit of perceived maximum efficiencies, in the process superseding human agency with technological agency and raising issues over autonomy and control. This reinforces the perspective of technical devices unsettling and shaking up practice (e.g. Hayles, 2017, p. 120), in the way that technical devices with agency can exceed human agency and consequently transform the nature of established human practice. This then raises a series of questions concerning the importance of understanding the relationship between digital devices and human agents and highlights several areas for future research in digital archaeology.

8.1 Authority in Practice

First among these is the question of where an appropriate balance between humans and technology should be struck in relation to archaeological practice. How much control is ceded to the digital device, and how much authority is retained by the human? This can be characterized as the human-in-the-loop problem (e.g. Broussard, 2018, p. 177ff) or the algorithm-in-the-loop problem (e.g. Green & Chen, 2019), but the key question is where in the loop does the human or algorithm sit? Where does control and authority lie?

The anthropocentric response is to claim that the human always remains in control, but the growing development of autonomous devices shows that this is not necessarily the case. Indeed, Amoore suggests that the human in the loop is an impossible figure, unable to comprehend the complexity of the systems: "How does one speak against the grain of the single output generated from millions of potential parameters?" (Amoore, 2019, p. 154). So, although it may be the objective to retain human authority over a process, how is this possible (and ethical) if that person does not understand the system? Archaeologists are already accustomed to using black-boxed devices in their practice: digital cameras and digital survey instruments are primarily managed in terms of their inputs and outputs with relatively little concern shown for the intervening technologies. On the other hand, archaeologists are also increasingly aware of the way that GIS functions are often used without a full grasp of the underlying algorithms (e.g. Brouwer Burg, 2017), a practice which is open to ethical challenge (e.g. Dennis, 2020, p. 212).

8.2 Automation in Practice

Secondly, and similarly a question of balance, is the role of automation within archaeological practice. Is Barceló's image of a specialized automated archaeologist capable of learning through experience to associate archaeological observations with explanations and using them to solve archaeological problems (Barceló, 2009, p. 352) a desirable one? How should human archaeologists treat the actions, decisions, and interpretations of digital automata? What is the status of an interpretation drawn by algorithms? Would there be a shift in practice from the study of past human culture towards the study of the digital past? Rather than pursuing automation and maximizing technological agency, is it more appropriate to perceive automation in terms of augmentation rather than replacement, where the digital devices assist and support, and archaeological agency remains unambiguously in human hands? Digital augmentation can be seen as extending human memory, processing power, predictive capacity, and visualization, for example, while the human contributes their ingenuity, imagination, flexibility, objective-setting, and contextualizing, etc., in a relationship that is more complementary than competitive. Digital devices can thereby allow archaeologists the opportunity to focus on identifying problems, defining approaches, and evaluating solutions, as well as providing relief from the mundane, manually intensive tasks (c.f. Tyukin et al., 2018, Section 4). In this light, the digital device enhances human capacity rather than reduces it, and agency is shared while authority lies firmly within the human realm. However, drift over time may subtly shift this balance. For example, digital augmentation may ultimately constrain the choices and outcomes presented to users (e.g. Bauer & Dubljević, 2020, p. 307; Danaher, 2018, p. 640), and such limitations, as a consequence of opacity, may be clandestine or even coercive in nature.

8.3 Computation in Practice

A third question relates to the effects of digitalization and digital agency on the practice of archaeological interpretation and analysis. For example, Parisi identifies a shift in reasoning through the introduction of cognitive computational devices, moving from "deductive truths applied to small data to the inductive retrieval and recombination of infinite data volumes" (Parisi, 2019, p. 92). The growth in big data and algorithmic approaches within archaeology has been similarly associated with a shift in theory and method (e.g. Huggett, 2020b, pp. S13–S15); one which switches the focus from a classic hypothesis-and-test approach to one of automated pattern-identification and a search for correlations, with the human expert reactive rather than proactive. To this end, machine learning methods incorporate uncertainty in the form of weighted probabilities, although the derivation of these is itself uncertain, and these probabilities are taken and reduced to a conclusion with a value between 0 and 1:

> … a single value distilled from a teeming multiplicity of potential pathways, values, weightings and thresholds. It is this process of condensation and reduction to one from many that allows algorithmic decision systems to retain doubt within computation and yet to place the decision beyond doubt (Amoore, 2019, p. 154).

What are the consequences for archaeological knowledge of practices which entail an arms-length relationship with the data, automated data processing and the identification of pattern, and the compression of degrees of uncertainty into a single conclusion in a system where the algorithms are poorly understood and the processes largely opaque? To what extent, therefore, may archaeological knowledge and understanding be shaped by invisible technological agency? How do computational digital agents change the relationship between archaeologists and past material culture?

9 Digital Agency and Responsibility

These questions give rise to a more fundamental issue: where does the responsibility lie in archaeological practice where the digital devices used possess varying degrees of agency and autonomy? It may seem self-evident, even reassuring, to suggest that responsibility should lie with the human rather than technological component, but the digital complexities of the technical solutions make this a problematic stance. This is essentially an ethical question. Dennis, for example, argues that archaeologists bear the ethical responsibility for the selection and use of digital tools:

> The use of a digital tool that cannot be understood by the user, or a digital method whose analytic processes cannot be explained by the user, is an inherently unethical choice (Dennis, 2020, p. 215).

Given the complexities of opacity, interpretability and comprehensibility discussed above, this seems to present a significant challenge. Is it realistic for all archaeologists using a particular digital tool to have this depth of knowledge rather than simply an awareness of the inputs, outputs, the limitations of the tool, and the appropriate modes of application? As Colley emphasizes, archaeology is embedded within a wider sociopolitical environment which begs the question as to who sets the ethical agendas, benefits from them, and can afford to act according to them (Colley, 2015, p. 16), and acknowledgment of this implies that the onus will not necessarily always be archaeological.

For example, a form of shared ethical responsibility may be recognized (c.f. Adam, 2008; Floridi, 2015, p. 134ff, for example), with the archaeological end user only the final link in an ethical chain of responsibility. That chain links the designers and developers of the technical device, the data collectors and aggregators, the device itself, and the end users. In turn, this implies that there are a host of potential ethical agents associated with a digital device, including the device itself: each will frequently have no knowledge of others in the chain, and most will be oblivious to any archaeological application, other than the end user archaeologist. Using these devices is therefore a largely unseen collaborative venture, combining human and technological agencies in a web of ethical relationships, practices, decisions, compromises, and solutions (Huggett, 2017, Section 7). Furthermore, the device itself will inherit ethical behaviors from the intentions of its designers and developers through constraints embedded in its data and software. For example, Ajunwa (2020, pp. 2–3) identifies the problems of "bias in, bias out" through incorrect training of algorithmic models to reflect human bias, and "data-laundering" where data processes produce biased results whilst appearing to be correct. Both represent ethical as well as technical problems. Whether an agential digital device can legitimately be seen as an ethical agent in its own right, rather than simply having implicit ethical agency inherited from its programmers (e.g. Moor, 2006, pp. 19–20), remains an open question, but its position within a chain of ethical responsibility remains regardless. Placing the burden of ethical responsibility entirely on the end user in a situation of distributed ethical agency and shared responsibilities therefore appears an over-reach, although that is not to say that the end user should not be appropriately informed and responsible for the eventual application.

10 Conclusions

What seems clear is that archaeologists cannot devolve ethical responsibility, authority, or control onto their devices without fundamentally changing the nature of archaeological practice. Moving the dial

towards greater digital agency will profoundly alter human practice. Some of these changes may seem positive, such as a reduction in the mundane, tedious, and repetitious in favor of more interesting, analytical, interpretative work, but others are more likely to be negative, not least those associated with the inscrutability and incomprehensibility of the devices in question.

Turkle describes the change in the relationship between humans and the digital as a shift from projection onto an object to engagement with an object (Turkle, 2005, p. 293). This shift can be perceived within the history of computer applications in archaeology, and the growth of digital agency and digital autonomy within archaeological practice requires that engagement to be appropriately informed, skeptical, and critical. As Hayles argues, we need to become:

> ... knowledgeable about how the interpenetrations of human and technical cognitions work at specific sites, and how such analyses can be used to identify inflection points, which ... emerge in interaction/intra-action ... to create new trajectories for the assemblages, providing more open, just, sustainable futures ... (Hayles, 2017, pp. 204–205).

This requires a commitment to understanding the nature of archaeological practice as part of what Hayles defines as "cognitive assemblages" (2017, p. 115): the complex interactions that exist between archaeologists, their digital tools, and the material objects of their study. In the process of examining these cognitive assemblages, unrecognized aspects of the human elements of archaeological practice are also revealed, situated alongside the digital components. As the future of archaeological practice will undoubtedly see a greater incorporation of digital devices possessing different degrees of agency and autonomy, the questions surrounding the human/technology relationship will need to remain a focus of near-constant negotiation.

Acknowledgements: A version of this paper was initially presented at the COST Arkwork conference "On Shifting Grounds: the study of archaeological practices in a changing world," held at the University of Crete in October 2019 and I am grateful in particular to Åsa Berggren, Antonia Davidovic-Walther, Dorina Moullou, and Rajna Šošić-Klindžić. Paul Reilly, Matt Edgeworth, and Costis Dallas provided valuable feedback in conversation, and I am also grateful to the two anonymous referees for their extensive and constructive responses. As ever, any errors or misconceptions are my own.

Conflict of interest: Author states no conflict of interest.

References

Adam, A. (2008). Ethics for things. *Ethics and Information Technology*, *10*(2–3), 149–154. doi: 10.1007/s10676-008-9169-3.

Ajunwa, I. (2020). The "black box" at work. *Big Data & Society*, *7*(2), 1–6. doi: 10.1177/2053951720938093.

Amoore, L. (2019). Doubt and the algorithm: On the partial accounts of machine learning. *Theory, Culture & Society*, *36*(6), 147–169. doi: 10.1177/0263276419851846.

Applin, S. A., & Fischer, M. D. (2015). New technologies and mixed-use convergence: How humans and algorithms are adapting to each other. *2015 IEEE International Symposium on Technology and Society* (*ISTAS*), (pp. 1–6). Dublin, Ireland: IEEE. doi: 10.1109/ISTAS.2015.7439436.

Atkinson, P. (2010). *Computer*. London: Reaktion.

Barad, K. M. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham: Duke University Press.

Barceló, J. A. (2009). *Computational intelligence in archaeology*. Hershey, PA: Information Science Reference.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. doi: 10.1016/j.inffus.2019.12.012.

Barrett, J. C. (2014). The material constitution of humanness. *Archaeological Dialogues*, *21*(1), 65–74. doi: 10.1017/S1380203814000105.

Bauer, W. A., & Dubljević, V. (2020). AI assistants and the paradox of internal automaticity. *Neuroethics*, *13*(3), 303–310. doi: 10.1007/s12152-019-09423-6.

Bennett, R., Cowley, D., & De Laet, V. (2014). The data explosion: Tackling the taboo of automatic feature recognition in airborne survey data. *Antiquity*, *88*(341), 896–905. doi: 10.1017/S0003598X00050766.

Bogost, I. (2015). The cathedral of computation. *The Atlantic*. Retrieved from: https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/.

Bonacchi, C., Bevan, A., Pett, D., Keinan-Schoonbaert, A., Sparks, R., Wexler, J., & Wilkin, N. (2014). Crowd-sourced archaeological research: The MicroPasts project. *Archaeology International*, *17*(0), 61–68. doi: 10.5334/ai.1705.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. Cambridge, Massachusetts: The MIT Press.

Brouwer Burg, M. (2017). It must be right, GIS told me so! Questioning the infallibility of GIS as a methodological tool. *Journal of Archaeological Science*, *84*, 115–120. doi: 10.1016/j.jas.2017.05.010.

Bucher, T. (2016). Neither black nor box: Ways of knowing algorithms. In S. Kubitschko & A. Kaun (Eds.), *Innovative Methods in Media and Communication Research* (pp. 81–98). Cham: Springer International Publishing. doi: 10.1007/978-3-319-40700-5_5.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. doi: 10.1177/2053951715622512.

Caraher, W. (2016). Slow archaeology: Technology, efficiency, and archaeological work. In E. W. Averett, J. M. Gordon, & D. B. Counts (Eds.), *Mobilizing the past for a digital future: The potential of digital archaeology* (pp. 421–441). Grand Forks, ND: The Digital Press at the University of North Dakota.

Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, *49*, 897–918. doi: 10.1007/s11186-020-09411-3.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. doi: 10.1093/analys/58.1.7.

Colley, S. (2015). Ethics and digital heritage. In T. Ireland & J. Schofield (Eds.), *The ethics of cultural heritage* (pp. 13–32). New York, NY: Springer New York. doi: 10.1007/978-1-4939-1649-8_2.

Cooper, M. A., & Richards, J. D. (Eds.). (1985). *Current issues in archaeological computing*. Oxford, UK: British Archaeological Reports.

Damilati, K., & Vavouranakis, G. (2021). What future for archaeology's past? In M. J. Boyd & R. C. P. Doonan (Eds.), *Far from Equilibrium: An archaeology of energy, life and humanity. A response to the archaeology of John C. Barrett* (pp. 115–129). Oxford, UK: Oxbow Books.

Danaher, J. (2018). Toward an ethics of AI assistants: An initial framework. *Philosophy & Technology*, *31*(4), 629–653. doi: 10.1007/s13347-018-0317-3.

Davis, D. S. (2020). Defining what we study: The contribution of machine automation in archaeological research. *Digital Applications in Archaeology and Cultural Heritage*, *18*, e00152. doi: 10.1016/j.daach.2020.e00152.

Demetis, D., & Lee, A. S. (2018). When humans using the IT artifact becomes IT using the human artifact. *Journal of the Association for Information Systems*, *19*(10), 929–952. doi: 10.17705/1jais.00514.

Dennis, L. M. (2020). Digital archaeological ethics: Successes and failures in disciplinary attention. *Journal of Computer Applications in Archaeology*, *3*(1), 210–218. doi: 10.5334/jcaa.24.

Dertouzos, M. L. (1997). *What will be: How the new world of information will change our lives*. London: Piatkus.

Dobres, M.-A. (2020). Agency in archaeological theory. In C. Smith (Ed.), *Encyclopedia of global archaeology* (pp. 75–82). Cham: Springer International Publishing. doi: 10.1007/978-3-030-30018-0_252.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv:1702.08608* [*Cs, Stat*], 1–13. Retrieved from: http://arxiv.org/abs/1702.08608

Ekbia, H. R., & Nardi, B. A. (2017). *Heteromation, and other stories of computing and capitalism*. Cambridge, MA: MIT Press.

Floridi, L. (2015). *The ethics of information*. Oxford: Oxford University Press.

Gell, A. (1998). *Art and agency: An anthropological theory*. Oxford, UK: Clarendon Press.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th international conference on Data Science and Advanced Analytics* (*DSAA*) (pp. 80–89). Turin, Italy: IEEE. doi: 10.1109/DSAA.2018.00018.

Gleicher, M. (2016). A framework for considering comprehensibility in modeling. *Big Data*, *4*(2), 75–88. doi: 10.1089/big.2016.0007.

Green, B., & Chen, Y. (2019). The principles and limits of Algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, *3*, 50. doi: 10.1145/3359152.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 93. doi: 10.1145/3236009.

Hanson, F. A. (2014). Which came first, the doer or the deed? In P. Kroes & P. -P. Verbeek (Eds.), *The moral status of technical artefacts* (pp. 55–73). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-007-7914-3_4.

Haraway, D. J. (1991). *Simians, cyborgs, and women: The reinvention of nature*. New York: Routledge.

Hayles, N. K. (2005). *My mother was a computer: Digital subjects and literary texts*. Chicago, IL: University of Chicago Press.

Hayles, N. K. (2017). *Unthought: The power of the cognitive nonconscious*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226447919.001.0001.

Hill, R. K. (2018). Assessing responsibility for program output. *Communications of the ACM*, *61*(8), 12–13. doi: 10.1145/3231166.

Hodder, I. (2012). *Entangled: An archaeology of the relationships between humans and things*. Malden, MA: Wiley-Blackwell.

Hodder, I. (2015). The asymmetries of symmetrical archaeology. *Journal of Contemporary Archaeology*, *1*(2), 228–230. doi: 10.1558/jca.v1i2.26674.

Horst, H. A., & Miller, D. (Eds.). (2012). *Digital anthropology*. London, UK: Berg.
Huggett, J. (2000). Computers and archaeological culture change. In G. Lock & K. Brown (Eds.), *On the theory and practice of archaeological computing* (pp. 5–22). Oxford, UK: Oxbow Books.
Huggett, J. (2012). Lost in information? Ways of knowing and modes of representation in e-archaeology. *World Archaeology*, *44*(4), 538–552. doi: 10.1080/00438243.2012.736274.
Huggett, J. (2017). The apparatus of digital archaeology. *Internet Archaeology*, *44*. doi: 10.11141/ia.44.7.
Huggett, J. (2020a). Capturing the silences in digital archaeological knowledge. *Information*, *11*(5), 278. doi: 10.3390/info11050278.
Huggett, J. (2020b). Is big digital data different? Towards a new archaeological paradigm. *Journal of Field Archaeology*, *45*(supp. 1), S8–17. doi: 10.1080/00934690.2020.1713281.
Humphreys, P. (2011). Computational science and its effects. In M. Carrier & A. Nordmann (Eds.), *Science in the context of application* (pp. 131–142). Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-9051-5_9.
Huvila, I., & Huggett, J. (2018). Archaeological practices, knowledge work and digitalisation. *Journal of Computer Applications in Archaeology*, *1*(1), 88–100. doi: 10.5334/jcaa.6.
Itkin, B., Wolf, L., & Dershowitz, N. (2019). Computational ceramicology. *ArXiv:1911.09960* [*Cs, Eess*], 1–20. Retrieved from: http://arxiv.org/abs/1911.09960
Jones, H. (2018). Geoff Hinton dismissed the need for explainable AI: 8 experts explain why he's wrong. *Forbes*. Retrieved from https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/
Kaufmann, M., & Jeandesboz, J. (2017). Politics and 'the digital': From singularity to specificity. *European Journal of Social Theory*, *20*(3), 309–328. doi: 10.1177/1368431016677976.
Knappett, C., & Malafouris, L. (Eds.). (2008). *Material agency: Towards a non-anthropocentric approach*. Boston, MA: Springer US. doi: 10.1007/978-0-387-74711-8.
Krakauer, D. (2016). Will A.I. Harm Us? Better to ask how we'll Reckon with our hybrid nature. *Nautilus*. Retrieved from: http://nautil.us/blog/will-ai-harm-us-better-to-ask-how-well-reckon-with-our-hybrid-nature
Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, Mass: Harvard University Press.
Leighton, M. (2015). Excavation methodologies and labour as epistemic concerns in the practice of archaeology. Comparing examples from British and Andean archaeology. *Archaeological Dialogues*, *22*(1), 65–88. doi: 10.1017/S1380203815000100.
Lindstrøm, T. C. (2015). Agency 'in itself'. A discussion of inanimate, animal and human agency. *Archaeological Dialogues*, *22*(2), 207–238. doi: 10.1017/S1380203815000264.
Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. doi: 10.1145/3233231.
Makridis, M., & Daras, P. (2012). Automatic classification of archaeological pottery sherds. *Journal on Computing and Cultural Heritage*, *5*(4), 1–21. doi: 10.1145/2399180.2399183.
Malafouris, L. (2015). Metaplasticity and the primacy of material engagement. *Time and Mind*, *8*(4), 351–371. doi: 10.1080/1751696X.2015.1111564.
Marwick, B. (2017). Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory*, *24*(2), 424–450. doi: 10.1007/s10816-015-9272-9.
McVicar, J. B., & Stoddart, S. (1986). Computerising an archaeological excavation: The human factors. In S. Laflin (Ed.), *Computer Applications in Archaeology 1986* (pp. 225–227). Birmingham, UK: Computing and Computer Science, University of Birmingham. Retrieved from: https://proceedings.caaconference.org/paper/16_mcvicar_stoddart_caa_1986/
Menary, R. (Ed.). (2010). *The extended mind*. Cambridge, Mass: MIT Press.
Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, *21*(4), 18–21. doi: 10.1109/MIS.2006.80.
Morgan, C. (2019). Avatars, monsters, and machines: A cyborg archaeology. *European Journal of Archaeology*, *22*(3), 324–337. doi: 10.1017/eaa.2019.22.
Olsen, B. (2010). *In defense of things: Archaeology and the ontology of objects*. Lanham [Md.]: AltaMira Press.
Olsen, B., & Witmore, C. (2015). Archaeology, symmetry and the ontology of things. A response to critics. *Archaeological Dialogues*, *22*(2), 187–197. doi: 10.1017/S1380203815000240.
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Allen Lane.
Onsrud, H., & Campbell, J. (2020). Being human in an algorithmically controlled world. *International Journal of Humanities and Arts Computing*, *14*(1–2), 235–252. doi: 10.3366/ijhac.2020.0254.
Parisi, L. (2019). Critical computation: Digital automata and general artificial thinking. *Theory, Culture & Society*, *36*(2), 89–121. doi: 10.1177/0263276418818889.
Plog, F., & Carlson, D. L. (1989). Computer applications for the All American Pipeline Project. *Antiquity*, *63*(239), 258–267. doi: 10.1017/S0003598X00075979.
Powlesland, D. (2016). 3Di – Enhancing the record, extending the returns, 3D imaging from free range photography and its application during excavation. In H. Kamermans, W. de Neef, C. Piccoli, A. G. Poluschny, & R. Scopigno (Eds.), *The three*

dimensions of archaeology (*Proceedings of the XVII UISPP world congress* (*1–7 September 2014, Burgos, Spain*) (pp. 13–32). Oxford, UK: Archaeopress.

Rains, M. J. (2015). Integrating database design and use into recording methodologies. In R. Chapman & A. Wylie (Eds.), *Material Evidence: Learning from Archaeological Practice* (pp. 79–91). Abingdon, Oxon: Routledge.

Rammert, W. (2012). Distributed agency and advanced technology. Or: How to analyze constellations of collective inter-agency. In J.-H. Passoth, B. Peuker, & M. Schillmeier (Eds.), *Agency without actors? New approaches to collective action* (pp. 89–112). London, UK: Routledge.

Reiner, P. B., & Nagel, S. K. (2017). Technologies of the extended mind: Defining the issues. In J. Illes (Ed.), *Neuroethics: Anticipating the future* (pp. 108–122). Oxford, UK: Oxford University Press.

Ribeiro, A. (2016). Against object agency. A counterreaction to Sørensen's 'Hammers and nails'. *Archaeological Dialogues*, *23*(2), 229–235. doi: 10.1017/S1380203816000246.

Ridge, M. (2013). From tagging to theorizing: Deepening engagement with cultural heritage through crowdsourcing. *Curator: The Museum Journal*, *56*(4), 435–450. doi: 10.1111/cura.12046.

Robb, J. (2010). Beyond agency. *World Archaeology*, *42*(4), 493–520. doi: 10.1080/00438243.2010.520856.

Robb, J. (2015). What do things want? Object design as a middle range theory of material culture: Object design as a middle range theory of material culture. *Archeological Papers of the American Anthropological Association*, *26*(1), 166–180. doi: 10.1111/apaa.12069.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (*Winter 2019*). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2019/entries/agency/

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 205395171773810. doi: 10.1177/2053951717738104.

Seaver, N. (2018). What should an anthropology of algorithms do? *Cultural Anthropology*, *33*(3), 375–385. doi: 10.14506/ca33.3.04.

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, *87*, 1085–1139. doi: 10.2139/ssrn.3126971.

Sørensen, T. F. (2016). Hammers and nails. A response to Lindstrøm and to Olsen and Witmore. *Archaeological Dialogues*, *23*(1), 115–127. doi: 10.1017/S1380203816000106.

Striphas, T. (2015). Algorithmic culture: *European Journal of Cultural Studies*, 18(40–5), 395–412. (Sage UK: London, England). doi: 10.1177/1367549415577392.

Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). Cambridge, Mass: The MIT Press. doi: 10.7551/mitpress/9780262014038.003.0009.

Taylor, J., Issavi, J., Berggren, Å., Lukas, D., Mazzucato, C., Tung, B., & Dell'Unto, N. (2018). 'The Rise of the Machine': The impact of digital tablet recording in the field at Çatalhöyük. *Internet Archaeology*, *47*. doi: 10.11141/ia.47.1.

Trier, Ø. D., Cowley, D. C., & Waldeland, A. U. (2019). Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, *26*(2), 165–175. doi: 10.1002/arp.1731.

Turkle, S. (1997). *Life on the screen: Identity in the age of the internet*. New York, NY: Simon and Schuster.

Turkle, S. (2005). *The second self: Computers and the human spirit*. Cambridge, Mass: MIT Press.

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science*, *18*(4), 347–361. doi: 10.1080/09540090600868912.

Tyukin, I., Sofeikov, K., Levesley, J., Gorban, A. N., Allison, P., & Cooper, N. J. (2018). Exploring automated pottery identification [Arch-I-Scan]. *Internet Archaeology*, *50*. doi: 10.11141/ia.50.11.

Verschoof-van der Vaart, W. B., & Lambers, K. (2019). Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology*, *2*(1), 31–40. doi: 10.5334/jcaa.32.

Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace* (pp. 11–22). Cleveland, OH: National Aeronautics and Space Administration. Retrieved from: https://archive.org/details/NASA_NTRS_Archive_19940022855/mode/2up

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, Mass: MIT Press.

Witmore, C. L. (2007). Symmetrical archaeology: Excerpts of a manifesto. *World Archaeology*, *39*(4), 546–562. doi: 10.1080/00438240701679411.

Woods, D., & Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical Issues in Ergonomics Science*, *1*(3), 272–282. doi: 10.1080/14639220110037452.

Wright, H., & Gattiglia, G. (2018). ArchAIDE: Archaeological automatic interpretation and documentation of ceramics. *Proceedings of the workshop on cultural informatics research and applications* (pp. 60–65). Nicosia, Cyprus.

Yang, K., Qinami, K., Fei-Fei, L., Deng, J., & Russakovsky, O. (2020). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (*FAT* 20*) (pp. 547–558). New York, NY: Association for Computing Machinery. doi: 10.1145/3351095.3375709.

Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, *41*(1), 3–16. doi: 10.1177/0162243915608948.