Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-021-00611-z

How WEIRD is Cognitive Archaeology? Engaging with the Challenge of Cultural Variation and Sample Diversity

Anton Killin[1,2] · **Ross Pain**[2]

Accepted: 28 December 2021

Abstract

In their landmark 2010 paper, "The weirdest people in the world?", Henrich, Heine, and Norenzayan outlined a serious methodological problem for the psychological and behavioural sciences. Most of the studies produced in the field use people from Western, Educated, Industrialised, Rich and Democratic (WEIRD) societies, yet inferences are often drawn to the species as a whole. In drawing such inferences, researchers implicitly assume that either there is little variation across human populations, or that WEIRD populations are generally representative of the species. Yet neither of these assumptions is justified. In many psychological and behavioural domains, cultural variation begets cognitive variation, and WEIRD samples are recurrently shown to be outliers. In the years since the article was published, attention has focused on the implications this has for research on extant human populations. Here we extend those implications to the study of ancient *H. sapiens*, their hominin forebears, and cousin lineages. We assess a range of characteristic arguments and key studies in the cognitive archaeology literature, identifying issues stemming from the problem of sample diversity. We then look at how worrying the problem is, and consider some conditions under which inferences to ancient populations via cognitive models might be provisionally justified.

✉ Anton Killin
anton.killin@anu.edu.au

Ross Pain
ross.pain@anu.edu.au

1 Philosophy Programme, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

2 School of Philosophy and Evolution of Cultural Diversity Initiative, Australian National University, Acton, ACT, Australia

Published online: 31 January 2022

1 Prelude

In 2010 Henrich and colleagues argued that the behavioural sciences face a serious methodological issue: most of the results in the field are produced using participants from WEIRD populations, yet these results often fail to replicate in cross-cultural studies (Henrich et al. 2010). Moreover, cross-cultural research suggests that WEIRD people are often outliers with respect to many cognitive and behavioural traits. So it seems that inferences from culturally localised samples to species-wide psychological claims are unjustified. The take home lesson is that, when it comes to behavioural psychology, culture matters; and often matters profoundly. There is no prior guarantee that an effect will be reproduced beyond the group it is found in, nor that it is an inlier with respect to the broader population.[1] Therefore, generalise with care.

So far, so worrying. But we think there is more to be worried about. The bulk of scholarly attention in the wake of Henrich and colleagues' article has focused on the issue of synchronous generalisation; that is, inferring from WEIRD populations to the (extant) human species at large ('generalisations'). However, as we will show, in the field of human cognitive evolution we often make diachronic inferences; that is, inferences from modern WEIRD populations to populations of ancient humans, their hominin forebears, or cousin lineages ('past projections').[2] And this too should be worrying—after all, ancient humans were certainly non-WEIRD.

Our focus is the discipline of cognitive archaeology. Cognitive archaeologists attempt to reconstruct the cognitive and cultural lives of ancient humans, Neanderthals, and their hominin ancestors by studying their material traces. One aim of the discipline is to shed some light on the evolution of human cognition. However, artefacts alone are insufficient to link information about behaviour to underlying cognitive capacity: a mid-range theory is required (Currie 2018; Binford 1972). So cognitive archaeologists typically appeal—either implicitly or explicitly—to a theory or model from the cognitive/psychological/behavioural (henceforth 'cognitive') sciences; one which is thought to be independently plausible. Using this model, inferences are made from artefacts to cognitive capacity.

We begin by introducing the challenge to cognitive archaeology in a little more detail, and provide some motivating examples (Section 2). We then outline four case studies which exemplify the issue of cross-cultural sample diversity in cognitive archaeology (Section 3). Next we look at how worrying the issue is, and briefly consider some conditions under which inferences from contemporary samples—most often WEIRD humans—to ancient populations might be provisionally justified (Section 4). Throughout, our mantra will be that further cross-cultural testing of cognitive models stands to strengthen or undermine the cognitive archaeological

[1] Indeed, if much of the mind's machinery is installed through cultural learning—whether the product of cultural evolution (Heyes 2018) or gene-culture coevolution (Laland 2017; Sterelny 2012)—then naturally that machinery will vary across cultures.

[2] For ease of expression, throughout this article we restrict the term 'human' to *H. sapiens*, and use 'hominin' for all *Homo* species, usually in reference to ancestral or cousin lineages of *H. sapiens*.

inference provisioned by those models. Of necessity, the discussion in section 4 is schematic and suggestive of further lines of inquiry; our primary goal in this paper is to foreground the problem at hand. Finally, Section 5 summarises, and outlines some future directions for research.

2 Introducing the Problem

Cognitive archaeology is a pluralistic and interdisciplinary research cluster with no single method. Rather, cognitive archaeologists employ a range of inferential strategies in order to reconstruct cognitive and cultural aspects of our evolutionary past in the general manner described above (see also Currie and Killin 2019; Currie 2018; Pain 2019; Malafouris 2020). A fairly common mode of inference, however, is *minimum-capacity* inference. This strategy attempts to identify the minimum cognitive prerequisites required for the production of some artefact by applying a model from the cognitive sciences to an artefact's construction.[3] For instance, we might infer that the production of Acheulean or Levallois technologies requires operational thought (Wynn 1979) or long-term working memory (Wynn and Coolidge 2004). Cognitive archaeologists running this mode of inference have often utilised cognitive models based on WEIRD samples, and this is potentially problematic.[4] In many domains relevant to the interdisciplinary study of evolution and human behaviour, individuals from WEIRD societies are not representative of human beings in toto (Henrich et al. 2010; Henrich 2020; Apicella et al. 2020). And while extant foragers are not Pleistocene hominins, and do not form a unified, monolithic whole (Astuti and Bloch 2010), their socioecology plausibly resembles more closely that of the ancient *H. sapiens* foragers from whom all humans today are descended. For these reasons, cognitive archaeology would do well to seek out mid-range theories corroborated

[3] There are many other strategies. For example, Hiscock (2014) infers that Acheulean toolmakers maintained apprentice learning systems managed by expert individuals based not on an explicit model from the cognitive sciences but a task analysis of Acheulean technology. Moreover, in addition to cognitive models, cognitive archaeologists may also leverage comparative, palaeoneurological, and other data (Wynn 2016; Currie and Killin 2019; Killin and Pain 2021). Here we are focused on the epistemic licence for an inference from archaeological artefact to cognitive capacity provisioned by a cognitive model.

[4] We say 'potentially', for not all cognitive domains might be sensitive to the WEIRD problem (we discuss this in Section 4.2); neither is it inevitable that any differences will entail effect sizes large enough to skew conclusions drawn by cognitive archaeologists via WEIRD data; nor is it known how many inferences will be undermined by the use of WEIRD data. So we do not claim that by using WEIRD data, cognitive archaeology is or has been wrongheaded. Our claim is that the use of cognitive models corroborated by more diverse samples stands to strengthen or undermine particular cognitive archaeological inferences, and, as we will argue, this is a point worth reflecting upon. Likewise, we note that it is for very good reasons that many cognitive/psychological/behavioural experiments are conducted on exclusively WEIRD subjects. The extent to which a study's sample is 'narrow' or 'diverse' is not a measure for distinguishing 'bad' from 'good' studies. The issue at hand is with inference beyond the sample. When generalisation is at issue, over-reliance on WEIRD participants is problematic for all the reasons detailed by Henrich et al. (2010), and much the same can be said of projection to the past. We thank an anonymous referee for pushing us on this.

by evidence from more diverse samples than the standard WEIRD range, and where possible to provide details of the sample so that inferences from WEIRD or otherwise narrow samples can be seen for what they are.

We do not have to look far for motivating cases. In their contribution to a special issue on the archaeology of children and childcare, Lew-Levy et al. (2020) argue that the results of developmental psychology experiments researching the innovation of children in WEIRD societies—results which suggest that children are poor innovators (e.g. Lister et al. 2020)—are not generalisable to children of small-scale, forager societies. And seemingly with good reason. Sterelny (2021) sums it thus:

> "...forager children are likely to innovate much more than WEIRD children, given that they grow up in an environment in which they have a lot of autonomy to play, to experiment and to explore the affordances of the material substrates to which they have access. In many forager cultures, experimental learning for oneself is positively encouraged. Furthermore, they routinely engage in peer-peer learning. While adults support social learning, they do relatively little directive teaching. Instead, they scaffold learning with equipment and raw materials, they provide occasional advice and they allow children to involve themselves with adult activities. Finally, children, even quite young children, do not just imitate in play adult economic and social activities, they practice those activities. They engage in subsistence activities, often in distinctive ways." (Sterelny 2021, p. 5)

The general concern should be clear: if the results of innovation studies do not generalise beyond WEIRD populations in extant humans, then we have reason to question their applicability to ancient humans. Furthermore, if we think that the learning environments of ancient populations were more akin to those of small-scale forager societies, then we should be prepared to treat samples from WEIRD populations as outliers.

Of course, innovation is typically assumed to be heavily influenced by sociocultural processes. So the preceding example is perhaps not particularly striking. It might be thought, however, that more 'base-level' processes—such as those involved in visual perception—would not vary much across contemporary human populations. Henrich et al. (2010) provide a motivating case that undermines this line of thinking—the Müller-Lyer illusion—which is now a classic example for establishing cross-cultural variability (Segal and Campbell 1966).[5] The two lines in the illusion are of equal length, though (to us and our colleagues, at least) they do not appear to be equal (see Fig. 1). Line A appears to us to be shorter. Researchers have tested the strength of the illusion by asking subjects how much longer Line A needs to be

[5] For review, see Phillips (2011). There is some debate about the replicability of the original results and the extent to which the cultural variation exists. For example, Deręgowski (2013) argues that only Müller-Lyer figures with interfin angles between 90° and 180° range are appropriate for testing cultural variation. Modified versions of the experiment continue to demonstrate cultural variation (e.g. Nestor 2017), however. While we acknowledge that the strength of the variation may be less than initially thought, we follow Henrich et al. (2010) in utilising this phenomenon as a motivating example, one which still demonstrates cultural variation in what we think is a surprising domain.

than Line B in order for the two lines to appear equal (see Fig. 2). This quantity varies widely across societies, along a continuum. At one end, WEIRD undergraduates require Line A to be about 20% longer than Line B; at the other end, the difference in line length San foragers required was indistinguishable from zero. For the San, there is no illusion. Much has been suggested about the causal role one's environment plays in bringing the illusion into effect (carpentered corners being ubiquitous in WEIRD societies and absent in the Kalahari; see McCauley and Henrich 2006; Henrich 2008).[6] However, a causal explanation of such variation is not what is at stake here. Rather, the case illustrates that even base-level visual processing can be subject to cultural variation. And, if even visual perception can vary across populations, the range of cognitive/psychological processes that are sensitive to culture is plausibly very broad. Moreover, the fact that the magnitude of the effect is strongest in WEIRD subjects is of concern, as this demonstrates that WEIRD people are outliers. We cannot thus argue that despite cultural variation WEIRD subjects are generally representative of the population. And if the inference from WEIRD people to San peoples is not justified, then we should be likewise concerned about any inferences to ancient populations, for, plausibly, the socioecology of the San more closely resembles that of ancient humans than that of WEIRD populations.

There are other examples, not involving illusions. For instance, researchers once thought, based on their studies of university students at their home institutions, that humans exhibit a right-hemisphere bias in face-recognition processing (see Henrich 2020, pp. 3-7). However, we now know that the process of becoming literate rewires our neural circuitry. One effect of this is that face-recognition processing is shifted to the right hemisphere (Dehaene et al. 2010, 2015). So the generalisation from the all-literate sample to *H. sapiens* is undermined, as it appears that people from non-literate societies do not exhibit such a bias (and, as it happens, are better facial-recognisers than literate people). Given that literate societies are relatively new, from an evolutionary perspective it would be an obvious mistake to infer on the grounds of the initial studies that ancient humans exhibited a right-hemisphere bias in face-recognition processing.

The purpose of this article is to apply this general line of critique to cognitive archaeology. We aim to draw out some of the issues raised by the challenge of cultural variation and sample diversity within the context of the historical sciences.

3 Case Studies

In this section we outline four examples from work in cognitive archaeology where WEIRD sampling issues are at play. In the first two, we target specific inferences concerning the evolution of spatial cognition and long-term working memory. We then expand our scope to look at research programs as a whole; namely the affordances framework and neuroarchaeology.

[6] See Nisbet (2003) for a broader take on cultural effects on perception.

3.1 Case Study 1: Wynn & Piagetian Theory

Thomas Wynn is a cognitive archaeology trailblazer,[7] and one of the most cited representatives of the discipline (as of the time of writing: over 6,200 citations according to GoogleScholar). Whether one agrees with his arguments or not, Wynn has been incredibly influential in the development of the field. For this reason, we pay particular attention to his work. Of course, we must be highly selective in an article of this length; but it is not from some fringe corner of the field that we select our examples. Indeed, Wynn's application of Piagetian psychological theory to the stone tool record is one of the earliest examples of a cognitive archaeological argument. This demonstrates that the problem of sample diversity was inherited (and indeed acknowledged) by cognitive archaeology from the outset; Piagetian theory's failure to generalise is one of the reasons Wynn discarded it in later work (see Wynn 2016). Our other examples demonstrate that the problem persists in various guises in the contemporary literature.

Before beginning, we note that our comments are not intended to be an indictment of Wynn's work in general; there are many areas of his work that are not subject to WEIRD concerns. Wynn (1985), for instance, utilises comparative data on non-human primates, and in other work, Wynn considers the implications of gross brain morphology as gained from palaeoneurological data. Our analysis merely draws attention to specific instances from Wynn's oeuvre where sample diversity issues are at play.[8]

The work of twentieth century Swiss psychologist Jean Piaget has been incredibly influential in developmental psychology and beyond. Piaget proposed a sequential stage model for individual intellectual development known as "genetic epistemology" (Piaget 1972). The stages are invariant, each necessary for the next, and based on a structuralism that conceives of intelligence as governed by a set of operational principles.[9] In 1979's seminal "The intelligence of later Acheulean hominids", Wynn applied Piaget's genetic epistemology model to the evolution of hominin intelligence. Specifically, Wynn's goal was to identify the spatial cognition capacities of ancient hominins by interpreting the 300,000 year old stone tools from Isimila Prehistoric Site, Tanzania, in terms of the final stage of Piaget's model, 'operational thought'.

Each stage of Piaget's model presents a set of operational principles—characterising development from birth until adolescence—which regulate thought. The final stage of the model sees individuals between 11 and 16 years old develop abstract reasoning, logical thought, and metacognitive reasoning resources. Moreover, according to the theory, *reversibility* and *conservation* are the two fundamental regulatory operators of these cognitive principles. Reasoning with the principle of transitivity exemplifies conservation. Schematically, if $x=z$ because $x=y$ and $y=z$ then it is due to the conservation of some property; an inference from

[7] As is well known to the fields' initiates: see, e.g., contributions to Overmann and Coolidge (2019).

[8] Thanks to an anonymous referee for pressing this.

[9] Although abstractions, the principles are conceived as reducible to (physical) brain activity.

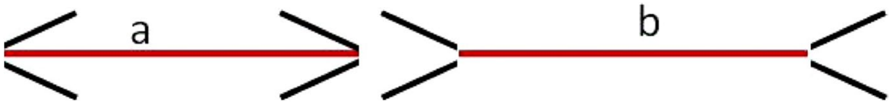

Fig. 1 The Müller-Lyer illusion. Source: Henrich et al. (2010), reproduced here with the permission of Cambridge University Press

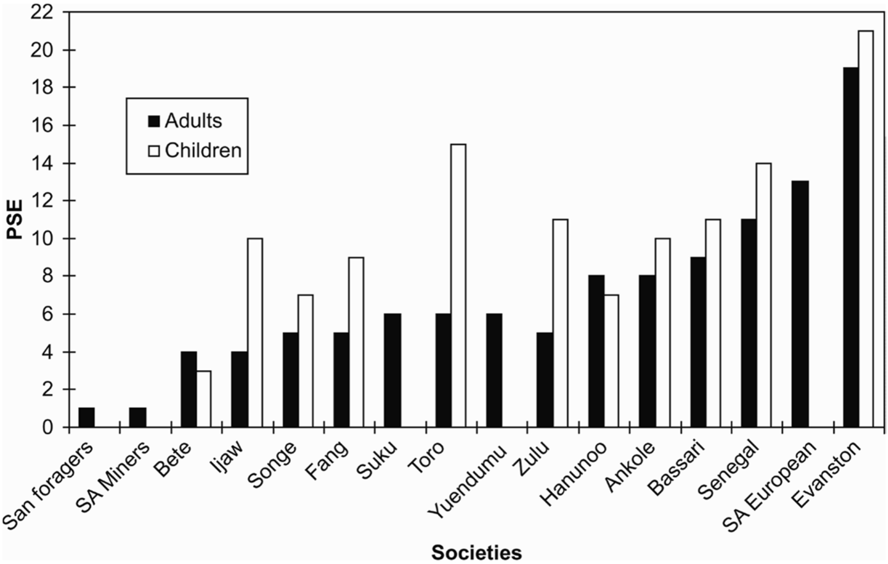

Fig. 2 Results from Segal and Campbell (1966). The point of subjective equality (PSE) indicates the percentage increase required for Line A to be judged as equal to Line B. Source: Henrich et al. (2010), reproduced here with the permission of Cambridge University Press

$x=y$ and $y=z$ to the conclusion $x=z$ is based on the same principle. Reversibility is the operation that takes one, through inversion, back to the starting point (e.g., $0+n\text{-}n=0$) or, through reciprocity, to an equivalence (e.g., $p \geq q$ and $p \leq q$ lead to $p=q$). Although the lines of reasoning here can be formalised with symbols, they are intended to capture operations which, at least most of the time, we employ informally in casual settings. (For further examples of these operations—including their relevance to classifications of kinship systems—see Wynn's paper.)

Wynn's interpretation of the stone tools from Isimila identified the targets of these psychological concepts in the tools' production. In doing so, he ran a minimum-capacity inference:

> "In order to manufacture all but the most rudimentary stone tools, however, flake removals must be related to one another in a fashion yielding the appropriate configuration or pattern. If a stone artefact presents a pattern of flake removals that could only have been organised by means of reversibility and/or conservation, then it must be concluded that the maker possessed

operational intelligence. I will show that the later Acheulean artefacts from the Isimila Prehistoric Site present such patterns." (Wynn 1979, p. 374)

His argument ran as follows. The reduction sequence from raw material to finished tool required the toolmaker to apply four kinds of operational spatial constructs: whole-part relations, qualitative displacement, spatio-temporal substitution and symmetry. Because each of these operations requires conservation and reversibility, the final stage of Piaget's model can be located in the makers of these tools. The bulk of Wynn's paper, then, is dedicated to explaining how one can infer the four spatial constructs from the knapping method identified in the Isimila bifaces, and how these relate to Piaget's two core ingredients of operational thought. (Although the details are interesting and informative, we must skip over them here; see Wynn 1979; also Wynn 1985 for discussion.)

Wynn's specific conclusions aside, our concern is with the use of Piaget's model as mid-range theory. Did it provide epistemic licence to Wynn's projection to the past? There are many reasons that Piagetian theory has gone out of fashion (Bjorklund 1997, 2018). According to Parker and Gibson (1979, p. 400) its evolutionary application is "Lamarckian and vitalistic"; and for other researchers, its reliance on the principle of recapitulation is "dangerous" (Renfrew 1982, p. 14). Importantly for our purposes though, Piaget's developmental model has long been criticised for its limited sample size: it was based on the observations of middle-class European children, and yet proposed as a general model of human cognitive development. Furthermore, as argued by Lancy (2010) and Shweder (2010), the general applicability of Piaget's model has been thoroughly undermined by cross-cultural research. Mead (1932) demonstrated that Piaget's developmental model (in its 1929 form) fails to generalise given her studies of an Admiralty Islands small-scale society, and Luria (1976) described alternative patterns of reasoning in Uzbek peasants in Central Asia. Wynn himself notes that a desiderata of a mid-range theory is that it is cross-culturally confirmed (Wynn 1985, p. 33), yet Piaget's theory fails this test. And if the theory cannot generalise from middle-class European children to children from the Admiralty Islands, then we should be skeptical of inferences made via the theory regarding the makers of the tools found at Ismilia. Moreover, we should be even more skeptical of any generalisations from those individual tool producers to other contemporaneous hominins, which Piaget's theory—and the scope of his conception of ontogeny as a singular, invariant process—appears to allow (cf. Wynn 1979, p. 383ff). It's no surprise that in later work Wynn discontinued this line of reasoning, eschewing the commitment to Piaget. It is important to note that this does not mean that the individuals who produced the Ismilia tools *did not* have the spatial constructs Wynn identifies; just that Piaget's cognitive model isn't the way to licence that inference. Wynn himself recognised this, and in later work he turned to comparative and palaeoneurological data, as well as a broad range of cognitive theories and frameworks via his subsequent teaming with psychologist Frederick Coolidge.

Wynn's early work is lauded by many cognitive archaeologists (e.g., Mithen 1996; Stade and Gamble 2019; McGrew et al. 2019; Davidson 2019), even though its hardline Piagetian justification would have few—if any—advocates today, as it helped to pave the way for a more mature discipline (and see Wynn 2016, pp. 8-10, for retroactive reflection on his Piagetian days). The case demonstrates that the problem of sample diversity in psychology was present at the beginnings of the cognitive archaeological project, and that its negative effects were acknowledged. The next subsection details a much more promising inference to the long-term working memory capacities of Neanderthals (Wynn and Coolidge 2004) by way of contemporary psychological and cognitive anthropological models.

3.2 Case study 2: Neanderthals and Long-term Working Memory

Recall that a minimum-capacity inference takes an artefact or tool industry—say, the Levallois technocomplex, associated with Neanderthals—and, via a cognitive theory, reaches a claim about the production of the trace—say, that the producers of those prepared-core technologies had the capacity for advanced (essentially, modern or very near-modern) long-term working memory. This is the claim defended by Wynn and Coolidge (2004).[10]

Levallois reduction is a task comprising multiple steps: "The first prepares a core with two distinct but related surfaces, one, a more convex platform surface that will include the striking platform, and a second flatter production surface from which the blank or blanks will be removed. The second step prepares the striking platform itself in relation to the axis of the intended blank. The third step is the removal, by hard hammer, of the blank or blanks." (Wynn and Coolidge 2004, pp. 473-474; see Figures 3 and 4 for a schematic representation and a photograph). When one blank is prepared, the method is called *preferential* and when two or more are prepared it is called *recurrent.* The recurrent method contains unidirectional, bidirectional and centripedal variants, each requiring different knapping techniques and platform preparation. Consequently, the 'end products' (e.g. Levallois points) are the result of not a single action-sequence but a plurality of operational schemas/methods (Boëda 1995).

Now, what of the mid-range theory through which the cognitive claim above is epistemically licenced? In this case, Wynn and Coolidge appeal to two cognitive models, one from cognitive neuropsychology (Ericsson and Kintsch 1995; Ericsson and Delaney 1999; Ericsson et al. 2000) and the other from cognitive anthropology (Keller and Keller 1996). Here we will focus on the latter.

Keller and Keller use the practice of smithing as their primary case study. According to their model, there are three crucial aspects to skilled expertise. First, 'the stock of knowledge', which is the information pool a smith acquires, builds, and maintains over many years of experience. This would include semantic/symbolic information, but also sensory information: visual, sonic, and tactile 'images' of procedures, materials, and so on. Second, an 'umbrella plan'; a mental model or representation of the end product intended as well as the tasks and subtasks required

[10] See Pain (2019) for further analysis of this line of reasoning. The conception of working memory comes from the work of Baddeley (e.g. Baddeley and Hitch 1974; Baddeley and Logie 1999; Baddeley 2001). Wynn and Coolidge also discuss how other aspects of the Neanderthal record complement this inference, so their argument is not solely based on Levallois.

for its production. Again, this would include both semantic/symbolic information and visual, sonic, and tactile information. Umbrella plans are similar to the concept of retrieval structures employed in cognitive neuropsychology: a given task recalls the relevant retrieval structure from long term memory, containing cues that facilitate encoding and retrieving associatively linked knowledge. Third, 'constellations', which are the requisite ideas and mental images, as well as the materials and tools, that enable each step of the process to be successfully accomplished or begun anew. Depending on the tasks involved, constellations can be deployed more or less automatically, or with full conscious attention. Active feedback between the smith and the constellation shapes the smith's actions and decisions in deploying the constellation.

Wynn and Coolidge claim that, in light of this model, Levallois reduction (and perhaps other Neanderthal technologies) demonstrates that the producers had long-term working memory capacities. They argue that, like smithing, prepared-core technology involves successfully completing a complex sequence of tasks requiring a stock of knowledge, an umbrella plan, and constellation, and that these concepts can be "directly applied" (p. 474). They say:

> "The sequence of actions that can be reconstructed for Levallois reduction resembles the sequence of action documented by the Kellers for blacksmithing: a sequential task with definable steps during which the artisan makes choices among a variety of specific techniques and procedures in order to complete each step and, ultimately, produce a finished product" (Wynn and Coolidge 2004, p. 474).

As is further outlined by Pain (2019), Wynn and Coolidge provide good reason to think the models' concepts map nicely onto aspects of the target phenomenon (the Levallois), so the conceptual framework is at least plausible.[11] Nonetheless, smithing may well be more culturally variable than the Kellers and Wynn and Coolidge implicitly assume, potentially restricting the scope of the model. Smithing, of course, is not an exclusively WEIRD activity. So cross-cultural research stands to strengthen or undermine the plausibility of the model's general application. As such, the cognitive capacities involved in the production of Levallois technologies may be distinct from that of modern smithing, undermining the analogy. This raises an important question: are all cases of artisan expertise organised in broadly similar ways, cognitively speaking? The concepts might appear to 'fit', but that fitting might yet overstate the long-term working memory of those ancient knappers (some interpretations from ecological psychology, and more broadly anti-representational approaches, would be consistent with this line of reasoning). So, even if the Keller's

[11] What's less plausible is Wynn and Coolidge's subsequent species-wide generalisation from the individual producers of the Levallois to Neanderthals broadly, e.g., "We believe that Neandertals regularly employed L-TWM [long-term working memory] that was as effective as any documented for modern humans" (Wynn and Coolidge 2004, p. 473). It is important to distinguish between a past projection and a species-wide generalisation from that projection. For even if some Neanderthals demonstrated a capacity (a producer of some tool) that does not imply that the capacity is distributed universally through the population/species.

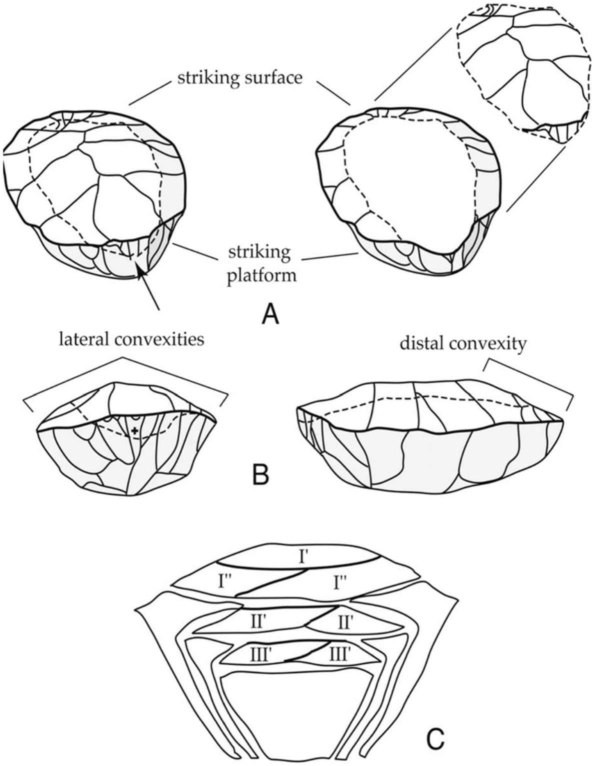

Fig. 3 Stages in prepared-core tool production. Source: Ambrose (2001), reproduced here with the permission of the American Association for the Advancement of Science

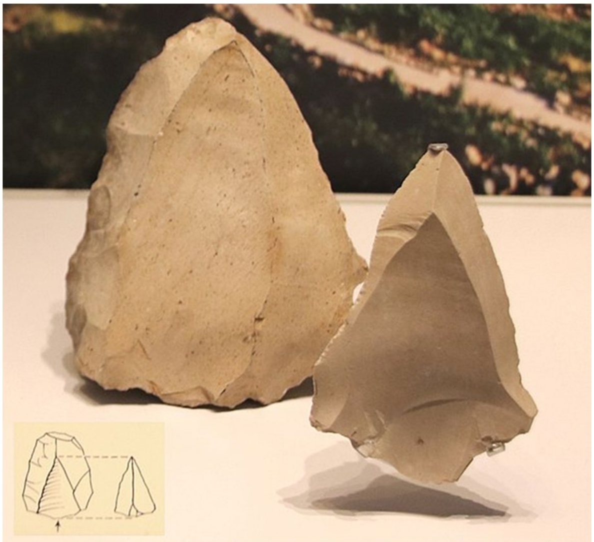

Fig. 4 Levallois stone tools from Tabun Cave (Mousterian culture) 50-250 kya. Reproducible under the Creative Commons CC0 1.0 Universal Public Domain Dedication. URL: https://commons.wikimedia.org/wiki/File:Production_of_points_%26_spearheads_from_a_flint_stone_core,_Levallois_technique,_Mousterian_Culture,_Tabun_Cave,_250,000-50,000_BP_(detail).jpg

model is applicable to skilled expertise in particular domains, it not only might fail to generalise, but also might not be suitable for projection to the deep past (after all, the Levallois dates as far back as roughly 300,000 years ago, and the last common ancestor of Neanderthals and *H. sapiens* lived roughly half a million years ago). Wynn and Coolidge admittedly are aware of this problem: "But are the cognitive underpinnings the same, that is, did Neandertal stone knappers have stocks of knowledge, make umbrella plans, and use constellations?" (p. 474). In their efforts to epistemically licence their inference, Wynn and Coolidge turn to modern skilled knappers, and they cite the work of Boëda (1995) and Baumler (1995). But these papers only describe the techniques involved in Levallois reduction, not cognitive underpinnings. Of course, it is relevant to take into account tool morphology, assemblage analysis and the constraints of knapping processes.[12] But testing the cognitive model, ideally utilising a diverse sample of participants, is also required.

Of course, all too often cognitive models based on culturally diverse samples are simply not available—precisely because of the bias towards WEIRD samples in the cognitive and behavioural sciences. Furthermore, for the most part, it is not the role of cognitive archaeologists to develop such models. Cognitive archaeologists are typically consumers, not producers of such models. But clearly, all things being equal, models licenced by data from culturally diverse samples should be preferred over those that are not. And where such models are not yet available, this could be acknowledged. The same point can be made about anthropological data: it is likely to be anthropologists, not cognitive archaeologists, who would provide rich descriptions, for example, of African indigenous smithing. Nonetheless, where available, such data stands to strengthen or undermine the epistemic licence for inferences such as Wynn and Coolidge's above.[13]

This example, like the Piagetian one preceding it, demonstrates that, in relying on models from the cognitive sciences, minimum-capacity inferences are at risk of inheriting the problem of cultural variation and sample diversity. However, perhaps unlike the Piagetian one, here we have an example of an argument that looks prima facie plausible (Pain 2019) before considering that challenge. In sum: further testing of the cognitive model using culturally diverse samples stands to strengthen or undermine the projection to the past.

3.3 Case Study 3: Affordances

The previous two examples targeted specific inferences in cognitive archaeology, focusing on the work of Wynn. In the following two case studies, we broaden our

[12] For all we have said, Wynn and Coolidge's conclusion may be right. Our point here is that it is too quick to assume as much based on their argument. Smithing may evince cultural variation in cognition; skilled expertise may be more cognitively heterogenous (within and/or across cultures) than assumed here—or may even necessitate a more representation-light explanation than the Kellers' model allows. These possibilities are empirically tractable and testable.

[13] Thanks to an anonymous referee for pressing these points.

scope to look at the challenge of cross-cultural variation and sample diversity to research programs more generally.

The dominant theoretical paradigm in contemporary cognitive science is representationalism/computationalism. Over the last 30 or so years, this dominance has come under increasing pressure from advocates of so-called 4E (embodied, enactive, extended and embedded) approaches. However, the situation in cognitive archaeology is somewhat different. This is partly due to a "founder effect" (Mayr 1942). Two archaeologist/psychologist duos were particularly influential in the development of cognitive archaeology: Coolidge and Wynn (as previously discussed), and Noble and Davidson (see e.g. Noble and Davidson 1996). And while the former duo work primarily in a representational framework, Noble and Davidson appeal to Gibsonian ecological psychology (see Davidson 2019 for a retrospective on the debate between these two). So frameworks that reject internalist, representation-focused accounts of cognition have been, and continue to be, a mainstay of cognitive archaeology (e.g., Malafouris 2020; Overmann 2017; see papers in DeMarrais et al. 2004; indeed, even Wynn (2020) has begun to engage with affordance theory).

So far our discussion has focused on Wynn's early- and mid-career work, which means it has been confined to representational/computational approaches to cognitive archaeology. In this section we want to broaden it to 4E approaches. We will focus on affordance theory, as this was Noble and Davidson's chosen framework. We will not target specific cognitive archaeological inferences here.[14] Rather, we want to track some work in the development of affordance theory, and show that the concept in general is likely to target phenomena which are culturally influenced. Consequently, care must be taken when inferring from WEIRD samples to ancient populations when appealing to the framework.

Affordances are understood as products of interactions between an organism and its environment. As different organisms have different body types, the same environment will produce different affordances for different organisms. For instance, to use an oft-quoted example: the wall in my office does not offer me the affordance of walking (I cannot walk on walls).[15] This is not the case for the spider that lives in the corner of my office (spiders can walk on walls). Walls (and ceilings) offer spiders the affordance of walking. Gibson (1979) argued that organisms perceive affordances directly, and the theory has subsequently become popular with those attempting to challenge the representational paradigm (e.g. Chemero 2009). Initial experimental work with human subjects focused on biomechanical models, where affordances were treated as a ratio between some aspect of a subject's body scale and some feature of the environment; for instance, between leg height and the height of a stair, or between eye height and the width of a gap (see e.g. Mark 1987; Burton 1992, 1994; Mark et al. 1999). A range of experiments appeared to generate reliable results. Most famously, Warren found a consistent ratio between leg height and stair height with regards to the boundary between stairs deemed "climbable" and stairs

[14] Though see, e.g., Silva-Gago et al. (2021), for a recent example of cognitive archaeological research via affordance theory.

[15] Shifting to single-author pronouns here for ease of expression.

deemed “unclimbable” (Warren 1984). Similar effects have been found in the case of eye height and the width of a gap deemed “crossable” (Jiang and Mark 1994). This work appears to lend weight to the body-scale account of affordances, and a body-scale account is one that would plausibly generalise across human populations.

However, as Chemero (2009, p. 143) notes, the property of organisms we are really interested in when it comes to affordances is their *ability* to perform some action. Aspects of body scale, such as leg height, are just a proxy for our ability to climb stairs that is easily quantifiable. This idea was tested by Cesari et al. (2003). Warren’s original stair climbing experiment used US college students as subjects, so the participants were mainly younger adults. By contrast, Cesari et al. used subjects with a range of different ages, including elderly people. They found that the ratio between leg height and stair height when it comes to the transition between climbable and unclimbable stairs varied considerably according to age. For elderly people, this ratio was considerably lower than the rest of the participants. Rather, flexibility seemed to also be playing an important role. Similar results have been produced in the case of gap crossing tasks (Chemero 2009, p. 145). Other work has contrasted subjects’ *judgments* of their ability to cross a gap and gap length with leg length and gap length (Chemero 2009). The former ratio was found to be much more highly correlated with the maximum gap judged crossable than the latter. Taken together, these results suggest that it is ability, not body scale, which is relevant in the perception of affordances.

All this is important for our purposes when we take into consideration two points. First, body-scale is a biological category, whereas ability is influenced by a range of biological and cultural factors. Stair climbing ability is a product of properties like leg height and flexibility, but it is also a learnt skill, and hence, to some extent, a product of culture. Furthermore, stair climbing ability will be to some extent determined by the prevalence of stairs in the environment—again, a cultural property. So populations that do not inhabit built environments will have different stair climbing abilities to those that do. The bottom line is that, if affordances are abilities, then they are influenced by culture. The situation here is roughly analogous to the case of the Müller-Lyer illusion (see section 2). There a ‘base-level’ perceptual effect was shown to be a product of architectural environments. Something similar is true of the trajectory of experimental work in affordances: an effect once thought to be more strictly confined to the biological domain is now understood as, to some extent, culturally determined. Second, the cited research above was carried out, not merely on WEIRD subjects, but, for the most part, on US college students.[16] And US colleges tend to be stair-rich environments.[17] Together, these two points suggest that any kind of generalised claim about affordances (beyond, perhaps, their existence) will need to be based on cross-cultural samples. And this is no different in the case of cognitive archaeology’s use of affordance theory.

[16] Details are as follows. Warren (1984): 54 male US college students; Cesari et al. (2003): 39 US citizens; Jiang and Mark (1994): 56 US college students; Chemero et al. (2010): 26 US college students.

[17] Of course, there are many non-WEIRD cultures with stair-rich environments. Though this is much less true of small-scale forager societies.

3.4 Case Study 4: Experimental Neuroarchaeology

Recently, cognitive archaeologists have developed a new strategy for producing inferences to the past. As we have seen, traditional strategies—such as minimum-capacity inferences—rely on interpreting the archaeological record using a model from the cognitive sciences. In contrast, experimental neuroarchaeology takes modern human subjects and investigates their brain activity during knapping tasks using neuroimaging techniques. This technique inherits a different set of theoretical and methodological problems. First and foremost among these is that modern human subjects are not, for instance, *H. erectus*; so any argument produced by this strategy is one by homology. This means that the strength of the inference involved depends on the neural similarity between modern *H. sapiens* and (again, for example) *H. erectus*, and the value of that ratio is very difficult to ascertain.

We will put these concerns aside here. Instead, we want to focus on issues of sample diversity in this research. In many ways, neuroarchaeology suffers similar problems to those we have identified in more traditional cognitive archaeology. But there is an upside here: the experimental aspect of neuroarchaeology allows us to more directly test cultural impacts on cognitive and behavioural capacities. In turn, this further illustrates the importance of sample diversity.

The majority of work in neuroarchaeology has so far focused on tool-language co-evolutionary hypotheses. In particular, research has tried to identify if there is any neural overlap between toolmaking and language production (e.g. Stout et al. 2008; Putt 2019). For instance, in their 2008 study Stout and colleagues took three expert tool knappers and used fluorodeoxyglucose positron emission tomography to assess the areas of the brain co-opted by both Oldowan and late Acheulean tool production tasks (Stout et al. 2008). They found that, compared with Oldowan toolmaking, late Acheulean toolmaking produced increased brain activity in areas of the brain associated with language production (in particular, Broca's area). More recently, Shelby Putt and colleagues have expanded the scope of this work using functional near-infrared spectroscopy (Putt et al. 2017; see Putt 2019 for an overview). Putt and colleagues were particularly interested in investigating whether the mode of learning—either verbal or non-verbal—had any effect on the brain regions co-opted by Oldowan and late Acheulean tasks. Participants in the experiment were taught to knap using either spoken language or via visual aids alone. Their results indicate that the mode of learning has a significant impact on parts of the brain co-opted during Acheulean-style toolmaking. This suggests that modern day knappers may be rehearsing the verbal instructions they were exposed to during learning.

Importantly though, the participants in these experiments were all WEIRD, and indeed either university professors or college students from the United States.[18] And this is worrying, as the need for sample diversity does not stop at Henrich et al.'s level of explanation (behavioural sciences) but is also a priority for neuroscience (Chiao and Cheon 2010). According to Chiao (2009), 90% of peer-reviewed

[18] Details are as follows. Stout et al. (2008): 3 professional archaeologists; Putt et al. (2017): 31 participants, recruited via posted flyers (presumably around the campus of Indiana University, Bloomington).

neuroimaging research comes from Western countries. Yet even between Westerners and East Asians there appear to be neuroscientific-level instances of cultural variation:

> "Westerners engage brain regions associated with object processing to a greater extent relative to East Asians, who are less likely to focus on objects within a complex visual scene (Gutchess et al. 2006). Westerners show differences in medial prefrontal activity when thinking about themselves relative to close others, but East Asians do not (Zhu et al. 2007). Activations in frontal and parietal regions associated with attentional control show greater response when Westerners and East Asians are engaged in culturally preferred judgments (Hedden et al. 2008). Even evolutionarily ancient limbic regions, such as the human amygdala, respond preferentially to fearful faces of one's own cultural group (Chiao et al. 2008, [...]). Taken together, these findings show cultural differences in brain functioning across a wide variety of psychological domains and demonstrate the importance of comparing, rather than generalizing, between Westerners and East Asians at a neural level." (Chiao and Cheon 2010, p. 89).

Moreover, there is a dearth of neuroimaging work utilising individuals from small-scale populations. While fMRI machines are not easily transportable, EEG methods (for instance) are non-invasive and the technology far more mobile. Chiao and Cheon's call for more effort to investigate neuroscientific research questions via sampling more diverse populations thus looks well justified.[19] Meanwhile, in the case of neuroarchaeology, it is important that inferences licenced via culturally localised samples are treated with caution.

Indeed, even the trajectory of research from Stout to Putt illustrates this. Stout purported to show neural overlap between toolmaking and language. Putt observed that Stout's experiments did not control for the method of learning of the participants. In testing the difference between visually taught participants and verbally taught participants Putt demonstrated that a cultural force—learning—could influence the results of neuroscientific testing. This trajectory serves as a cautionary tale, yet also illustrates an important opportunity. As an experimental research program, neuroarchaeology can actually test cultural variables, even when operating with limited sample diversity—Putt's work shows that the way in which a skill is learnt affects the neural substrates it co-opts. This is not an option for traditional cognitive archaeology, which operates by interpreting the archaeological record through a cognitive model.

[19] Suppose this was done and the results were broadly similar: Acheulean-style but not Oldowan-style knapping engages language-processing areas of the brain across a very broad sample. This corroboration would add licence to past projection, in our view, rendering it more plausible than we'd otherwise grant based on a small and narrow sample.

4 How Worried Should We Be?

In the previous section, we outlined examples in which cognitive archaeologists use models, frameworks, or experimental research from the cognitive sciences which have been built using limited samples or are otherwise susceptible (or potentially susceptible) to the challenge from cultural variation. In this section, we look in more detail at how worrying this situation is. We begin with some conceptual work clarifying the issue of generalisation and sample diversity in the context of inference to the deep past. Then, drawing on Henrich et al. (2010), we look at some conditions under which generalisations from limited sample sizes might be justified.

4.1 Sample Diversity, Generalisation, and Inference to the Past

Recall the issue of sample diversity as it faces the cognitive sciences. The problem is that, by generalising from WEIRD samples, researchers assume that either there is no cross-cultural variation, or that WEIRD people are generally representative of the species. However, cross-cultural research has shown that in many cases these assumptions do not hold. So sample diversity is required to make reliable claims at the larger population- or species-level.

Now, take the case in which we use a cognitive model derived from WEIRD samples alone, and then apply that model to some artefact from the stone tool record using a minimum-capacity inference. What precisely is the worry here? In the first instance, notice that this does not *necessarily* undermine the inference to the cognitive capacity of the maker of the artefact. If the minimum-capacity inference holds—and that is a big 'if'—then we should be confident that the maker of the artefact at least had the relevant capacity. Remember, the lesson from Henrich et al. (2010) is *generalise* with care; at this stage in the process, no generalisation has been made. Though perhaps unlikely given the differences in lifeways between WEIRD people and past peoples, it is at least logically possible that we are successfully inferring from one culturally localised sample to another. A distinct problem arises, however, when we generalise from that minimum-capacity inference to a population-wide or species-wide claim in the deep past. That generalisation would be justified if we had reason to believe that there was no cultural variation in the population (which is perhaps only plausible for pre-*Homo* populations) or that WEIRD people were generally representative of the species and diachronically so, but of course this is not the case. If, however, the cognitive model at hand has been corroborated by cross-cultural research, then, all other things being equal, researchers have greater epistemic licence to extend the inference to that of the population- or species-level.

To illustrate this point, consider Wynn and Coolidge's minimum-capacity inference concerning the long-term working memory abilities of Neanderthals (section 3.2). The inference here used Keller and Keller's cognitive anthropological model of skilled expertise, based on their observations of smithing. In the absence of cross-cultural testing of that model, the minimum-capacity inference may well licence claims regarding the long-term working memory capacities of those individuals, and perhaps their wider cultural group, found in situ with Levallois technology (that is, if we are satisfied that the past projection via that cognitive model is provisionally licenced—we have discussed above some challenges for this). In other words, we may be satisfied that the inference takes us from the technology, via the cognitive model, to the cognitive capacity of the individuals who produced the technology. But even then we would not be justified in making species-wide claims regarding Neanderthals—that Neanderthals in general had modern or near-modern long-term working memory—as Wynn and Coolidge do (and then go on to draw additional inferences based on that generalisation, see Wynn and Coolidge 2004, pp. 478ff). On the other hand, if Keller and Keller's model was corroborated by cross-cultural research, all other things being equal, the inference would more reliably allow species-wide generalisation. At the very least, such inferences would gain credibility. The lesson is this: sample diversity allows us to make more plausible generalisations in the present, and this is important insofar as it provides greater epistemic licence for generalisation in the past.

4.2 Are There Cases Where Generalisations From Weird Samples are 'Safe'?

We have thus far argued that lack of sample diversity is inferentially problematic for cognitive archaeology, insofar as it utilises cognitive models licenced by data on WEIRD subjects, and outlined a range of cases—both specific and more general. In this section, we want to examine some cases where traditional minimum-capacity inferences might be more plausible. Henrich et al. (2010) suggest a range of conditions under which generalising from WEIRD samples to contemporary humans might be provisionally justified:

1. Effects found in "[...] cognitive domains related to attention, memory and perception…".
2. Effects "... measured at a physiological or genetic level".
3. "... generalisations from one well-studied universal phenomenon to another similar phenomenon".
4. Effects demonstrated "...in other species, such as rats or pigeons…".
5. Effects "...which are evident among infants".
6. Effects in brain regions "...less responsive to experience". (Henrich et al. 2010, p. 79)

We think this list can be reduced to two conditions: [a] where an effect is reliably thought to be a product of biology; and [b] where an effect is reliably thought to be a product of a cultural universal. 2 is a way of reaching the conclusion in [a], while 4, 5 and 6 are methods of eliminating culture as a potential cause of an effect, thereby increasing the likelihood that it is biological. 3 is one way of reaching the conclusion in [b]. Finally, the intuition driving 1 looks to be related to [a], but Henrich et al. note that the work of their paper "...does not bolster this intuition" (Henrich et al. 2010, p. 79). Simply put, there are two reasons why we might think a cognitive effect might be universal—it is either part of human biology or it is, for whatever reason, common to all human cultures. Are there areas of research to which cognitive archaeology contributes that satisfy these conditions? We consider cases from research on language and theory of mind, but begin with two general observations.

First, a distinct possibility is that satisfying [a] in the present may be insufficient for the production of reliable inferences to our deep past. This will occur as the morphology of hominin body shapes changes through phylogenetic space. Thus effects that generalise across modern day human populations due to their being measured at a physiological level will not apply as we move further back in time.

Second, it might be thought that long periods of cultural uniformity, such as the Acheulean, signal that either [a] or [b] has been satisfied.[20] This would only apply to those who commit to *cognitive* explanations of uniformity—those who posit demographic or environmental causes might not consider either to be satisfied (see Pain 2019 for an overview).

As a result of Chomsky's influence, the view that our ability to produce language is innate and domain specific (e.g. Berwick and Chomsky 2015; Chomsky 2007) is widespread. If we thought this view was true, then we might think that [a] would be satisfied. The evolution of language was a key focus of early cognitive archaeological studies (e.g., Gibson and Ingold 1993; Mithen 1996; Noble and Davidson 1996), so perhaps much of this work is immune from sample diversity concerns. However, there is increasing acknowledgement amongst researchers regarding the causal importance of culture in the evolution of language. This includes gene-culture coevolutionary accounts (e.g., Laland 2017; Tomasello 2005, 2010), and more radical accounts that deny biologically produced language-specific capacities in human ontogeny (e.g., Christiansen and Chater 2016; Heyes 2018). If these theories are on the right track for language, [a] would either not be satisfied at all, or look much more difficult to satisfy. In addition, recent tool-language co-evolutionary hypotheses look to develop accounts where the evolution of syntactical features of language involves co-opting existing capacities evolved to support toolmaking (e.g., Stout et al. 2008; Stout and Chaminade 2012; Planer and Sterelny 2021). Syntactical capacities are sometimes thought to be one of—if not *the*—biologically-endowed, domain-specific language capacity. These accounts, however, suggest that the phylogenetic ancestry of that capacity was heavily driven by culture. This again raises questions about the ability to satisfy [a].

'Theory of mind' or 'mindreading' refers to our ability to infer, understand, or simulate the mental states of another individual. For instance, one of us might infer from the yawns of an audience that participants in the lecture are bored. Our theory of mind capacities are embedded in a broader framework of orders of intentionality. These orders begin with the awareness of our own mental states, and progress from there. For instance: a lecturer intends (1); that their audience understand (2); that Stout believes (3); that Chomsky disagrees (4); with Tomasello's commitment to domain-general processes in theorising about language (5); and so on. Theory of

[20] There are those who question notions of uniformity during the Acheulean (e.g. Stout 2011). Naturally, people of this persuasion will not be particularly persuaded by this point.

mind is thus located in the second order of intentionality and beyond. Recently, Cole (e.g., Cole 2016; Cole 2019) has produced a range of studies attempting to correlate theory of mind capacities and orders of intentionality with the lithic record (see also Planer 2017). Now, one might think that the ability to interpret other people's mental states is a culturally produced capacity (Heyes 2018) and is perhaps also *distinctive* of human beings. Theorists of this persuasion may well think that theory of mind is something approaching a cultural universal, and hence that it satisfies [b]. However, Cole (2019) has argued that the record suggests more variation in orders of intentionality capacities than universal models indicate, which would undermine this conclusion.

We have run, very briefly, through two cases where one might have reason to think that [a] or [b] is satisfied—however, recent work tells against this conclusion in both cases. Of necessity we have given these research programs short shrift; a full analysis is beyond the scope of this article and is an avenue for future research.

5 Conclusion and Future Directions

Henrich et al. (2010) identified an important problem for the behavioural and cognitive sciences: the tension between the empirical reality of cultural variation and the narrow cultural representativeness of most study samples. Cultural variation in many domains engenders cognitive variation, and WEIRD participants are often not representative of the human species. We have argued that cognitive archaeology inherits this problem insofar as it uses models (and frameworks, experimental research, etc.) from the cognitive sciences to provide epistemic licence for its inferences to the past. Our examples demonstrate the breadth of the challenge to cognitive archaeology: it is not restricted to any specialisation or theoretical paradigm, and has historical roots.

We have claimed that rather than being methodologically flawed, inferences to cognitive capacity from a physical trace via a cognitive model stand to be strengthened or called into question by further testing of the cognitive model against more culturally diverse samples. Corroboration is possible, so the situation is not totally dire; there is cause for *optimism* (Currie 2018). We have also suggested that there may be some cases where the force of the problem is mitigated—hypothetically, some minimum-capacity inferences from particular cognitive models might be more or less plausible in light of *independent* reason to think an inference can be applied outside the sample, e.g., when an effect is reliably thought to be the product of biology, or when an effect is reliably thought to be the product of a cultural universal. That said, our brief analysis suggests that these conditions look difficult to satisfy. A full analysis is an important avenue of future research.[21]

[21] Another route to optimism utilises causal-association inferences, not discussed in this paper. See Currie and Killin (2019) and Pain (2019) for analysis; and Killin and Pain (2021) for the role of multiple independent lines of evidence converging on capacity in cognitive archaeological theorising.

Finally, as we have stressed throughout, theoretical resources from the cognitive sciences are an important part of the cognitive archaeologist's inferential toolkit. Yet cognitive science is in a state of live debate. This poses a problem for cognitive archaeologists: which principles should guide cognitive archaeologists when selecting a cognitive model for use as a mid-range theory? We (Killin and Pain 2021) have recently proposed two complementary, though non-exhaustive solutions: theory choice should be guided by consilience (convergence from multiple independent lines of evidence) and methodological pluralism. This article proposes a further, compatible principle. That is, all other things being equal, cognitive archaeologists should prefer models tested and corroborated by diverse samples; or, at the very least, where effort has been undertaken to assess how well the results (very often from WEIRD samples) might project to the deep past. Since cultural effects produce cognitive and behavioural variations even in contemporary populations, due caution must be taken when projecting inferences to ancient hominins, and it would be no bad thing to heed this caution in model selection. After all, the above considerations suggest that past projections require a developmental cognitive mid-range theory, one that takes into account the effects of culture on cognition. In turn, this implies that reconstructing ancient minds requires, to some degree at least, reconstructing their cultures.[22] This challenge too is suggestive of future research.

Acknowledgements For comments on previous versions of this work, we thank Kim Sterelny, Adrian Currie, Andra Meneganzin, Rachael Brown, and two anonymous referees. The article also benefited from feedback from audiences: an ANU School of Philosophy seminar, the "New Work in Social and Cultural Evolution" workshop (ANU, 2021), and the online ISHPSSB 2021 conference. We gratefully acknowledge support from the Australian National University. RP was supported by an ANU University Research Scholarship and the ANU Futures Scheme. AK thanks the National Library of New Zealand and Lilburn Trust for supporting his research, though his opinions don't necessarily reflect theirs.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

References

Ambrose, S. 2001. Paleolithic technology and human evolution. *Science* 291 (5509): 1748–1753.
Apicella, C., A. Norenzayan, and J. Henrich. 2020. Beyond WEIRD: a review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior* 41: 319–329.
Astuti, R., and M. Bloch. 2010. Why a theory of human nature cannot be based on the distinction between universality and variability: lessons from anthropology. *Behavioral and Brain Sciences* 33: 83–84.
Baddeley, A.D. 2001. Is working memory still working? *American Psychologist* 11: 851–864.

[22] Here we thank Kim Sterelny for helpful discussion.

Baddeley, A.D., and G.J. Hitch. 1974. Working memory. In *Recent advances in learning and motivation*, ed. G.A. Bower, 74–90. New York: Academic Press.
Baddeley, A.D., and R.H. Logie. 1999. Working memory: the multiple-component model. In *Models of working memory: mechanisms of active maintenance and executive control*, ed. A. Miyake and P. Shah, 28–61. New York: Cambridge University Press.
Baumler. 1995. Principles and properties of lithic core reduction: implications for Levallois technology. In *The definition and interpretation of Levallois technology*, ed. H.L. Dibble and O. Bar-Yosef, 3–19. Madison, Wisconsin: Prehistory Press.
Berwick, R.C., and N. Chomsky. 2015. *Why only us: language and evolution*. Cambridge, MA: MIT Press.
Binford, L. 1972. *An archaeological perspective*. New York: Seminar Press.
Bjorklund, D.F. 1997. In search of a metatheory for cognitive development (or, Piaget is dead and I don't feel so good myself). *Child Development* 68 (1): 144–148.
Bjorklund, D.F. 2018. A metatheory for cognitive development (or "Piaget is dead" revisited). *Child Development* 89 (6): 2288–2302.
Boëda, E. 1995. Levallois: a volumetric construction, methods, a technique. In *The definition and interpretation of Levallois technology*, ed. H.L. Dibble and O. Bar-Yosef, 41–68. Madison, Wisconsin: Prehistory Press.
Burton, G. 1992. Nonvisual judgment of the crossability of path gaps. *Journal of Experimental Psychology: Human Perception and Performance* 18 (3): 698–713.
Burton, G. 1994. Crossing without vision of path gaps. *Journal of Motor Behavior* 26 (2): 147–161.
Cesari, P., F. Formenti, and P. Olivato. 2003. A common perceptual parameter for stair climbing for children, young and old adults. *Human Movement Science* 22 (1): 111–124.
Chemero, A. 2009. *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
Chemero, A., C. Klein, and W. Cordeiro. 2010. Events as changes in the layout of affordances. *Ecological Psychology* 15 (1): 19–28.
Chiao, J.Y. 2009. Cultural neuroscience: a once and future discipline. *Progress in Brain Research* 178: 287–304.
Chiao, J.Y., and B.K. Cheon. 2010. The weirdest brains in the world. *Behavioral and Brain Sciences* 33: 88–90.
Chiao, J.Y., T. Iidaka, H.L. Gordon, J. Nogawa, M. Bar, E. Aminoff, N. Sadato, and N. Ambady. 2008. Cultural specificity in amygdala response to fear faces. *Journal of Cognitive Neuroscience* 20 (12): 2167–2174.
Chomsky, N. 2007. Biolinguistic explorations: design, development, evolution. *International Journal of Philosophical Studies* 15: 1–21.
Christiansen, M.H., and N. Chater. 2016. *Creating language: integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.
Cole, J. 2016. Accessing Hominin Cognition: Language and Social Signaling in the Lower to Middle Palaeolithic. In *Cognitive models in Palaeolithic archaeology*, ed. T. Wynn and F.L. Coolidge, 157–195. Oxford: Oxford University Press.
Cole, J. 2019. Knapping in the Dark: Stone Tools and a Theory of Mind. In *Squeezing Minds from Stones: Cognitive Archaeology and the Evolution of the Human Mind*, ed. K.A. Overmann and F.L. Coolidge, 355–375. Oxford: Oxford University Press.
Currie, A. 2018. *Rock, bone, ruin: an optimist's guide to the historical sciences*. MIT Press.
Currie, A., and A. Killin. 2019. From things to thinking: cognitive archaeology. *Mind & Language* 34 (2): 263–279.
Davidson, I. 2019. Evolution of cognitive archaeology through evolving cognitive systems: a chapter for Tom Wynn. In *Squeezing Minds from Stones: Cognitive Archaeology and the Evolution of the Human Mind*, ed. K.A. Overmann and F.L. Coolidge, 79–101. Oxford: Oxford University Press.
Dehaene, S., F. Pegado, L.W. Braga, P. Ventura, G. Nunes Filho, A. Jobert, G. Dehaene-Lambertz, R. Kolinsky, J. Morais, and L. Cohen. 2010. How learning to read changes the cortical networks for vision and language. *Science* 330: 1359–1364.
Dehaene, S., L. Cohen, J. Morais, and R. Kolinsky. 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews: Neuroscience* 16 (4): 234–244.
DeMarrais, L., C. Gosden, and C. Renfrew, eds. 2004. *Rethinking materiality: the engagement of mind with the material world*. McDonald Institute for Archaeological Research Press.
Deręgowski, J.B. 2013. On the Müller-Lyer illusion in the carpentered world. *Perception* 42 (7): 790–792.

Dibble, H.L. 1995. Biache Saint-Vaast, Level IIA: a comparison of analytic approaches. In *The definition and interpretation of Levallois technology*, ed. H.L. Dibble and O. Bar-Yosef, 93–116. Madison, Wisconsin: Prehistory Press.

Ericsson, K.A., and P.F. Delaney. 1999. Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In *Models of working memory: mechanisms of active maintenance and executive control*, ed. K.A. Ericsson and P.F. Delaney, 257–297. New York: Cambridge University Press.

Ericsson, K.A., and W. Kintsch. 1995. Long-term working memory. *Psychol Rev* 102 (2): 211–245.

Ericsson, K.A., V. Patel, and W. Kintsch. 2000. How experts' adaptations to representative task demands account for the expertise effect in memory recall: comment on Vicente and Wang (1998). *Psychol Rev* 107 (3): 578–592.

Gibson, J.J. 1979. *The ecological approach to visual perception*. Houghton and Mifflin.

Gibson, K.R., and T. Ingold. 1993. *Tools, language and cognition in human evolution*. Cambridge: Cambridge University Press.

Gutchess, A.H., R.C. Welsh, A. Boduroglu, and D.C. Park. 2006. Cultural differences in neural function associated with object processing. *Cognitive, Affective and Behavioral Neuroscience* 6: 102–109.

Hedden, T., S. Ketay, A. Aron, H.R. Markus, and J.D.E. Gabrieli. 2008. Cultural influences on neural substrates of attentional control. *Psychological Science* 19 (1): 12–17.

Henrich, J. 2008. A cultural species. In *Explaining culture scientifically*, ed. M. Brown, 184–210. University of Washington Press.

Henrich, J. 2020. *The WEIRDest people in the world: how the West became psychologically peculiar and particularly prosperous*. New York: Farrar, Straus & Giroux.

Henrich, J., S.J. Heine, and A. Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33: 61–135.

Heyes, C. 2018. *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press.

Hiscock, P. 2014. Learning in lithic landscapes: a reconsideration of the hominid toolmaking niche. *Biological Theory* 9 (1): 27–41.

Keller, C.M., and J.D. Keller. 1996. *Cognition and tool use: the blacksmith at work*. Cambridge: Cambridge University Press.

Killin, A., and R. Pain. 2021. Cognitive archaeology and the minimum necessary competence problem. *Biological Theory*. https://doi.org/10.1007/s13752-021-00378-7.

Laland, K. 2017. *Darwin's unfinished symphony: how culture made the human mind*. Princeton University Press.

Lancy, D.F. 2010. When nurture becomes nature: ethnocentrism in studies of human development. *Behavioral and Brain Sciences* 33: 99–100.

Lew-Levy, S., A. Milks, N. Lavi, S.M. Pope, and D.E. Friesem. 2020. Where innovations flourish: an ethnographic and archaeological overview of hunter-gatherer learning contexts. *Evolutionary Human Sciences* 2: e31. https://doi.org/10.1017/ehs.2020.35.

Lister CJ, Walker B, Fay N (2020) Innovation and enculturation in child communication: a cross-sectional study. *Evolutionary Human Sciences* 2: e56. https://doi.org/10.1017/ehs.2020.57

Luria, A.R. 1976. *Cognitive development: its cultural and social foundations*. Harvard University Press.

Jiang, Y., and L.S. Mark. 1994. The effect of gap depth on the perception of whether a gap is crossable. *Perception & Psychophysics* 56: 691–700.

Malafouris, L. 2020. How does thinking relate to toolmaking? *Adaptive Behavior*. https://doi.org/10.1177/1059712320950539.

Mark, L.S. 1987. Eyeheight-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance* 13 (3): 361–370. https://doi.org/10.1037/0096-1523.13.3.361.

Mark, L.S., Y. Jiang, S.S. King, and J. Paasche. 1999. The impact of visual exploration on judgments of whether a gap is crossable. *Journal of Experimental Psychology: Human Perception and Performance* 25 (1): 287–295. https://doi.org/10.1037/0096-1523.25.1.287.

Mayr, E. 1942. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press.

McCauley, R.N., and J. Henrich. 2006. Susceptibility to the Müller-Lyer illusion, theory neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology* 19 (1): 79–101.

McGrew, W.C., T. Falotico, M.D. Gumert, and E.B. Ottoni. 2019. A simian view of the Oldowan: reconstructing the evolutionary origins of human technology. In *Squeezing minds from stones: cognitive archaeology and the evolution of the human mind*, ed. K.A. Overmann and F.L. Coolidge, 13–41. Oxford: Oxford University Press.

Mead, M. 1932. An investigation of the thought of primitive children, with special reference to animism. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 62: 173–190.

Mithen, S. 1996. *The prehistory of the mind*. London: Thames & Hudson.

Nestor P (2017) Cross-cultural variations in the Muller-Lyer illusion. In: Cognitive Science Research Papers Winter 2017, pp. 158-163. Los Angeles: Department of Psychology, University of California. https://www.psych.ucla.edu/sites/default/files/documents/Winter%202017.pdf

Nisbet, R.E. 2003. *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. Nicholas Brealey.

Noble, W., and I. Davidson. 1996. *Human evolution, language and mind: a psychological and archaeological inquiry*. Cambridge, Melbourne: Cambridge University Press.

Overmann, K.A. 2017. Materiality and numerical cognition: a material engagement theory perspective. In *Cognitive models in Palaeolithic archaeology*, ed. T. Wynn and F.L. Coolidge. Oxford: Oxford University Press.

Overmann, K.A., and F.L. Coolidge, eds. 2019. *Squeezing minds from stones: cognitive archaeology and the evolution of the human mind*. Oxford: Oxford University Press.

Pain, R. 2019. What can the lithic record tell us about the evolution of hominin cognition? *Topoi*. https://doi.org/10.1007/s11245-019-09683-0.

Parker, S.T., and K.R. Gibson. 1979. A developmental model for the evolution of language and intelligence in early hominids. *Behavioural and Brain Sciences* 2: 367–408.

Phillips, W.L. 2011. Cross-cultural differences in visual perception of color, illusions, depth, and pictures. In *Cross-cultural psychology: contemporary themes and perspectives*, ed. K.D. Keith, 160–180. Malden: MA, Wiley-Blackwell.

Piaget J (1929) *The child's conception of the world* (Harcourt, Brace & Company).

Piaget, J. 1972. *The principles of genetic epistemology, trans. W Mays*. London/New York: Routledge.

Planer, R.J. 2017. Talking about tools: did early Pleistocene hominins have a protolanguage? *Biological Theory* 12 (4): 211–221.

Planer, R.J., and K. Sterelny. 2021. *From signal to symbol: the evolution of language*. Cambridge, MA: MIT Press.

Putt, S.S., S. Wijeakumar, R.G. Franciscus, and J.P. Spencer. 2017. The functional brain networks that underlie Early Stone Age tool manufacture. *Nature Human Behaviour* 1: 1–8.

Putt, S.S. 2019. The stories stones tell of language and its evolution. In *Squeezing minds from stones: cognitive archaeology and the evolution of the human mind*, ed. K.A. Overmann and F.L. Coolidge, 304–318. Oxford: Oxford University Press.

Renfrew, C. 1982. *Towards an archaeology of mind*. Cambridge: Cambridge University Press.

Segal MH, Campbell DT, Herskovits (1966) *The influence of culture on visual perception*. Bobbs-Merrill.

Shweder, R.A. 2010. Donald Campbell's doubt: cultural difference or failure of communication? *Behavioral and Brain Sciences* 33: 109–110.

Silva-Gago M, Fedato A, Hodgson T, Terradillos-Bernal M, Alonso-Alcalde R, Bruner E (2021) Visual attention reveals affordances during Lower Palaeolithic stone tool exploration. *Archaeological and Anthropological Sciences* 13: 145. https://doi.org/10.1007/s12520-021-01413-1

Stade, C., and J. Gamble. 2019. In three minds: extending cognitive archaeology with the social brain. In *Squeezing minds from stones: cognitive archaeology and the evolution of the human mind*, ed. K.A. Overmann and F.L. Coolidge, 319–331. Oxford: Oxford University Press.

Sterelny, K. 2012. *The evolved apprentice: how evolution made humans unique*. Cambridge, MA: MIT Press.

Sterelny, K. 2021. Veiled agency? Children, innovation, and the archaeological record. *Evolutionary Human Sciences*. https://doi.org/10.1017/ehs.2021.9.

Stout, D., and T. Chaminade. 2007. The evolutionary neuroscience of toolmaking. *Neuropsychologia* 45: 1091–1100.

Stout, D. 2011. Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 366 (1567): 1050–1059.

Stout, D., and T. Chaminade. 2012. Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society of London*. *Series B, Biological sciences* 367 (1585): 75–87. https://doi.org/10.1098/rstb.2011.0099.

Stout, D., N. Toth, K. Schick, and T. Chaminade. 2008. Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. *Philosophical Transactions Royal Society B: Biological Sciences* 363: 1939–1949.

Stout, D., R. Passingham, C. Frith, J. Apel, and T. Chaminade. 2011. Technology, expertise and social cognition in human evolution. *European Journal of Neuroscience* 33 (7): 1328–1338.

Stout, D., E. Hecht, N. Khreisheh, B. Bradley, and T. Chaminade. 2015. Cognitive demands of lower Paleolithic toolmaking. *PLoS ONE* 10 (4): e0121804.

Tomasello, M. 2005. *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. 2010. *Origins of human communication*. MIT Press.

Warren, W.H. 1984. Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance 10* (5): 683–703.

Wynn, T. 1979. The intelligence of later Acheulean hominids. *Man, New Series* 14 (3): 371–391.

Wynn, T. 1985. Piaget, stone tools, and the evolution of human intelligence. *World Archaeology* 17 (1): 32–43.

Wynn, T., and F.L. Coolidge. 2004. The expert Neandertal mind. *Journal of Human Evolution* 46: 467–487.

Wynn, T. 2016. Evolutionary cognitive archaeology. In *Cognitive models in Paleolithic archaeology*, ed. T. Wynn and F.L. Coolidge, 1–20. New York: Oxford University Press.

Wynn (2020) Ergonomic clusters and displaced affordances in early lithic technology. *Adaptive Behavior*, https://doi.org/10.1177/1059712320932333

Zhu, Y., L. Zhang, J. Fan, and S. Han. 2007. Neural basis of cultural influence on self representation. *Neuroimage* 34: 1310–1317.