UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

Department of Pure and Applied Sciences

Ph.D. PROGRAMME IN:

Research Methods in Science and Technology

CYCLE XXXVIII

FROM BONES TO BELIEFS: AN ANALYSIS OF EVIDENTIAL REASONING IN PALEOANTHROPOLOGY AND ARCHAEOLOGY

ACADEMIC DISCIPLINE: PHIL-02/A LOGIC AND PHILOSOPHY OF SCIENCE

Coordinator: Prof. Luca Lanci

Supervisor: Prof. Vincenzo Fano

Ph.D. student: Matteo Bedetti

ACADEMIC YEAR
2024/2025

He who understands baboon would do more towards metaphysics than Locke

Charles Darwin, 1838: Notebook M

ABSTRACT

This dissertation investigates evidential reasoning in paleoanthropology and Paleolithic archaeology. More specifically, it analyzes the inferential processes through which fragmentary traces, mostly bones, tools and sediments, are transformed into claims about our evolutionary past.

The first chapter examines the debate on *Homo naledi*'s alleged burials, with a particular emphasis on the notion of "evidential standards". It provides a characterization of this notion, and it is shown that in the case of *naledi,* the debate concerns not so much what these standards are, but how they should be interpreted. The second chapter addresses variability in Paleolithic burial practices, showing how this phenomenon has tangible consequences on how we define burials and how we identify them in the archaeological record. The third chapter analyzes the use of null hypotheses in cognitive archaeology, suggesting that importing statistical tools into non-experimental contexts produces heuristic nulls that suffer from problems of testability. Inference to the best explanation is offered as a better alternative. The final chapter explores the integration between evolutionary psychology and paleoanthropology, showing how the heuristics of hypothesis generation of the former should be informed and constrained by the data of the latter.

Overall, the thesis advances a philosophical account of how knowledge about human evolution is built from fragmentary traces, highlighting both the potentialities and limitations of this process. It concludes by outlining some directions for future inquiry.

Table of contents

INTRODUCTION
THE FORGOTTEN SCIENCES? PALEOANTHROPOLOGY, ARCHAEOLOGY AND THE PHILOSOPHY OF HISTORICAL SCIENCES

1. Man the Hunter (?)

When did humans begin hunting? Chimpanzees and bonobos, our closest living relatives, show a propensity for hunting small game, especially in open, seasonal environments. By phylogenetic continuity, this points to an early propensity for pursuit and ambush already present in the last common ancestor between *Pan* and humans. We can therefore state with some certainty that *Ardipithecus ramidus* also had the same predisposition for hunting. In these early stages of human evolution, however, we only have indirect evidence of meat consumption. The long archaeological silence comes to an end around 2.6 million years ago at Gona, where we find the first clear traces of butchery on medium and large animals. Later, at FLK Zinj in Olduvai Gorge, the picture becomes unmistakable: cut marks cluster on the areas richest in meat, limb bones are fractured for marrow, and mortality profiles show a preference for prime adults. Together, these traces point to primary access to carcasses rather than opportunistic scavenging. The adopted strategy is that of ambush, employing wooden spears that are carefully shaped, hardened, and used with considerable precision[1]. Although wood rarely survives deep time, the lithic record preserves its reflection in microwear consistent with woodworking; residues from resilient timbers on bifaces attest to plant processing; field experiments confirm that even untipped spears launched can penetrate thick hide. Living chimpanzees that sharpen sticks with their teeth to dispatch concealed prey show that the behavioral steps are well within reach. This is further corroborated by the evolution of postcranial morphology. In *Homo erectus* the configuration of the shoulders and torso suggests an enhanced capacity to store and release elastic energy, enabling accurate and forceful overhand throws. By ~1.8 Ma, our ancestors are skilled hunters of medium and large game. In this reconstruction, hunting is a key ecological strategy that reshaped many aspects of early hominin life, from diet and mobility to cooperation and food sharing, and ultimately influenced their evolutionary paths[2].

[1] The oldest wooden spears date back to around 400,000 years ago (the Schöningen spears; Thieme, 1997).

[2] The reconstruction just outlined draws on a series of studies by Bunn, Pickering, Domínguez-Rodrigo, and their collaborators (e.g., Bunn & Pickering, 2010; Pickering, 2013; Domínguez-Rodrigo & Pickering, 2017). According to this account, hominins gained early access to animal carcasses through ambush hunting. This is, however, not the only reconstruction of the evolution of our ancestors' hunting behavior. An alternative and widely discussed hypothesis proposes that early humans practiced persistence hunting, a strategy based on exhausting prey through long-distance

And yet the apparent certainty of this narrative rests on a net of material inferences: the position of cut marks and fracture patterns, mortality profiles of butchered bovids, residue and microwear analyses, comparisons with primate behavior, biomechanical studies of fossil bones, and many others. Each of these lines of evidence is individually fragile, based on long chains of inferences and a great deal of background knowledge from different disciplines. The majority of tools made by organic materials are lost and their presence can only be indirectly inferred in the record. Different interpretations of the same archaeological and paleontological findings abound. We speak as if we know exactly when and how man started hunting, but to what extent can we trust these reconstructions? Curiously, many researchers in the field of human evolution would say: not much. In the next section, we will turn to the views of these researchers.

2. Different shades of pessimism

It is noteworthy that, at times, paleoanthropologists, as well as scholars in related fields studying human evolution, adopt a pessimistic outlook on their discipline. It is possible to read extremely negative assessments, such as that paleoanthropology is «a caricature of a science» and «modern human origins research in particular, fails to meet the standards of a science like physics» (Clark, 2000, p. 852) or «no science can ever be historical» (Gee, 2000, p. 8), expressed by paleoanthropologists and paleontologists themselves. This pessimism should seem, in the first instance, puzzling. From new discoveries, theoretical advancements and inclusion of new techniques, "modern human origins research" hardly seems like a caricature of a real science[3].

Notwithstanding this, pessimism in paleoanthropology and related fields takes several forms. In this section, we aim to provide a preliminary taxonomy of some of its main varieties. Our focus will be on the views expressed by paleoanthropologists themselves, though, where relevant, we will also acknowledge contributions from researchers in other fields. This taxonomy is not intended to suggest that the different forms of pessimism are mutually exclusive (indeed, many of these are interrelated), nor that it exhausts the range of possible attitudes. Rather, it is meant to illustrate some of the characteristic ways in which pessimism has been articulated within the field.

pursuit (e.g., Bramble & Lieberman, 2004). Clearly, the two hypotheses are compatible: our ancestors may have hunted by a combination of ambush and persistence running.

[3] As evidence of this, during the three years of my PhD numerous new discoveries concerning human evolution have emerged. These include the earliest evidence for the structural use of wood around 480,000 years ago (Barham et al., 2023); indications that the capacity to adapt to arid environments was already present in *Homo erectus* (Mercader et al., 2025); proteomic analyses revealing that fossils attributed to *Homo longi* actually belong to Denisovan individuals (Fu et al., 2025), new evidence that some *sapiens* cultures practiced cannibalism (Margines et al., 2025); the application of a new dating method that has backdated the earliest rock art in Indonesia (Oktaviana et al., 2024) and many others.

"Every few years our knowledge of human evolution is turned upside down!" Although such remarks are rarely made explicitly in the literature, they are often voiced by those inside and outside the discipline in more or less informal settings. The sentiment behind them appears to be widespread, and it is among the first reactions I encounter when I tell academics from other fields that I work on the epistemology of paleoanthropology. It seems to reflect a common attitude among the educated public that in paleoanthropology there are no empirical acquisitions that are stable, and that our reconstruction of human evolution changes radically every few years. The discipline's great popularity is probably partly to blame. On the one hand, progress in this field is remarkably rapid; on the other, the media often resort to sensationalist headlines when reporting news about human evolution (Chapter 1; see also Delisle, 2018). As famously observed by paleoanthropologist Tim D. White: «[n]o suid skulls grace the covers of Nature or garner headlines like "new pig skull completely overturns all previous theories of pig evolution"» (1995, p. 369). This social prestige has other consequences for scientific practice: for example, given their value, fossil specimens are preserved with extreme jealousy and rarely shared physically with the scientific community Although this trend appears to be undergoing change: for instance, initiatives such as MorphoSource now provide open access to 3D digital data of fossil specimens and related materials.

Uniqueness. Smith and Wood (2016, 2017; Smith 2016) recently argued that paleoanthropology faces a specific epistemic challenge: *uniqueness*. A human trait is unique when there is no contrastive class against which it can be compared. Their main example is bipedality, a defining feature of human evolution. While bipedal locomotion can be compared to that of other species if we are interested in biomechanics, it becomes unique, and thus explanatorily problematic, when we ask *why* it evolved in our lineage (see also Cartmill, 2002). In such cases, the absence of comparable evolutionary trajectories limits empirical constraints and increases the risk of underdetermination. The implication, according to Smith and Wood, is that explanations of unique traits are ultimately untestable or unfalsifiable (they are «just-so stories» [Gould & Lewontin, 1979], according to Smith, 2016). While this does not render paleoanthropology unscientific per se, it still suggests a form of *pessimism*: if many key human traits, such as cultural abilities or large brain size, are unique in this sense, then the scope of scientific explanation in human evolution is drastically limited. Paleoanthropology would thus be confined to reconstructing morphological and behavioral changes, without explaining why they occurred[4].

[4] See Villmoare and Kimbel (2024) for an optimistic reply to the uniqueness challenge. According to them, the uniqueness of certain typical human traits has been exaggerated, and it is (almost) always possible to analyze a trait in a comparative approach (see also Buskell & Currie, 2021, for a philosophical perspective). See Smith and Wood (2024) for a reply.

Narrativity. Notoriously, Misia Landau argued that paleoanthropological reconstructions are not only scientific explanations but also narratives shaped by cultural motives. In *Narratives of Human Evolution* (1991), she showed how accounts of human origins often echo mythic structures, portraying the story of humankind as that of a hero who undergoes various trials to realize his true self—suggesting that their plausibility rests as much on narrative coherence as on empirical evidence. This raises the question of how far explanations in human evolution derive authority from data versus from culturally embedded narrative forms.

But there is also a much stronger sense in which reconstructions in paleoanthropology are defined as "narratives", which relates to the issue of uniqueness. Some hypotheses may be narrative in the sense of being overly speculative, constructing causal explanations *post hoc* to fit the available data rather than generating independent, testable predictions (see again Smith, 2016; see also Chapter 4)[5].

Stagnation and repetition. There is a sentiment that seems quite common among paleoanthropologists. Indeed, several have pointed out that their discipline suffers from a lack of progress, and that it often finds itself fossilized in the same, recurring debates (such as those between unilinear and multilinear phylogenetic hypotheses, or between the Out of Africa and Multiregional models). Paleoanthropology would be a field of research completely riddled with controversy, where almost every aspect is constantly questioned and there are few fixed points. (Delisle, 2012, presents the views of several paleoanthropologists who seem to be leaning in this direction). The reasons identified behind this stagnation can be varied. It could be due to a lack of good communication practices (Tattersall & Eldredge, 1977, p. 204), or other non-empirical factors such as clashing paradigms (e.g. Hammond & Stocking, 1988).

Pervasiveness of bias. Another concern is that the discipline would be particularly prone to biases and preconceptions. This proclivity has a series of consequences that naturally lead to pessimism. There are several expressions of this sentiment in literature, but for a recent case, see the debate related to *Homo naledi*'s burials and engravings. In the open discussion attached to one of *naledi* team's papers, one of the reviewers states[6]: «[u]ltimately, this manuscript presents evidence that those who are pro the deep antiquity of intentional mark-making by *Homo* (and

[5] See Currie and Sterelny (2017) for a defense of narratives in science. When understood in a nuanced way, narratives are not simply unconstrained speculations but can be difficult to formulate and serve very specific epistemic functions, such as guiding the search for new empirical evidence.

[6] We will have more to say about the publication process recently adopted by *eLife*, the scientific journal where the papers on *H. naledi*'s purported burials and engravings have been published, but for the moment it should suffice to say that it involves the publication of a pre-print with the reviews attached, an assessment by the journal and (sometimes) a reply by the authors. The latter can decide then whether or not to update the preprint in light of reviews received.

possibly even other genera) will find enough evidence to support; while those skeptical of such claims will find enough methodological flaws and evidential limits to refute those claims» (Berger et al., 2023, p. 29). The idea is that debates in paleoanthropology are driven less by empirical discovery than by theoretical preconceptions that shape the interpretation of evidence. This is, in fact, one of the factors often cited as a cause of the alleged circularity of paleoanthropology. As noted earlier, debates within the discipline are thought to recur cyclically because they are ultimately guided not by new evidence, but by researchers' preconceptions and theoretical commitments.

Failure to comply with evolutionary biology standards. Addressing questions about human evolution, paleoanthropology falls squarely within the domain of evolutionary biology. Nevertheless, some have argued that the discipline does not sufficiently adhere to the standards of evolutionary science, particularly in its taxonomic practices (see for example Schwartz, 2018; Tattersall, 1992).

No science at all/caricature of a science. A final concern is that paleoanthropology is not a science at all. Unlike Landau and Cartmill, who argue that paleoanthropology can incorporate narrative elements, or is entirely narrative if it merely proposes a chronology of events without linking them through causal regularities, some believe that this discipline does not qualify as a science in any meaningful sense. The reasons for this claim vary, but they often include explicit comparisons with physics, emphasizing that paleoanthropology does not meet the same standards of testability and falsifiability, and thus falls short of equivalent scientific rigor (see Clark's and Gee's quotes above).

These are just some of the ways in which pessimism about paleoanthropology manifests itself (see also Henke, 2015). As noted, this pessimism is at least curious, in light of the advances in the discipline. This raises the question of what philosophers of science have had to say on the matter.

3. The debate around the historical sciences

Until a few decades ago, historical sciences such as geology, paleontology, paleoanthropology and archaeology received little attention from philosophers of science. As a consequence, debates in philosophy of science had been flattened mainly on experimental sciences. With the exception of evolutionary biology, most philosophers of science have taken experimental sciences, and physics in particular, as a paradigmatic example of science. Carl Gustav Hempel was among the first to frame the philosophy of science in these terms (see especially Hempel, 1942). His covering law or deductive-nomological model, strongly inspired by the experimental sciences, physics in

particular, held that scientific explanations should take the form of logical deductions from general laws and initial conditions. By prioritizing generality and predictability, Hempel's framework implicitly marginalized disciplines that do not produce or rely on such laws. While well suited to physics and chemistry, it poorly captured the explanatory practices of historical sciences such as geology, paleontology, and archaeology. Hempel nonetheless maintained that explaining singular historical events was scientifically legitimate only if such explanations could be subsumed under general laws. This normative emphasis on lawlike regularities made historical explanations appear inferior or "less scientific", since "genetic" explanations, as Hempel called them, amounted merely to explanation sketches. Historical sciences were evaluated against the standards of the experimental sciences, using the latter as a benchmark.

This general disinterest in the historical sciences persisted throughout the rest of the 20th century, with a few rare exceptions. Among them are William H. Dray, who developed the concept of "how-possibly explanations" to capture the preliminary stages of explanation in the historical sciences (1968), Alison Wylie, whose works on the philosophy of archaeology (e.g., 1985) remain one of the most significant contributions in the field, and a few others. Despite these efforts, they did not give rise to a sustained or structured debate within philosophy of science.

Over the past two decades, however, this trend has reversed, and there is now a lively debate within the philosophy of science concerning the nature of the historical sciences (Chapman & Wylie, 2016; Cleland, 2002, 2011; Currie, 2018; Forber & Griffith, 2011; Jeffares, 2008; Tucker, 2011; Turner, 2005; 2007). The emergence of a structured debate on the historical sciences was sparked mainly by Carol Cleland's 2002 paper *Methodological and epistemic differences between historical science and experimental science*. The article framed the dichotomy between historical and experimental sciences and has since served as a starting point for numerous philosophical reflections on the peculiarities of the former. At first, the debate was not framed in terms of optimism or pessimism about the epistemic prospects of these disciplines. Yet, in examining the nature of historical research in comparison with experimental inquiry, philosophers often implicitly provided reasons either for or against epistemic pessimism. For example, Cleland claims that the present overdetermines the past, due to the time asymmetry of causation (2002, 2011). That is, events in the past leave widespread and diverse effects, and we do not need every trace left by an event to infer that that event happened. To re-utilize the example made by Cleland, you don't have to collect every shade of glass to infer that a window broke. This fact about the time asymmetry of causation underwrites the particular method employed by historical scientists to investigate the past: namely, common cause explanation. Historical scientists usually proceed by looking for a common cause in the past that explains the presence of different traces in the present.

But this fact underwrites the adoption of an optimist stance as well. If the present systematically overdetermines the past, it means that there is always more evidence in the world than is needed to reconstruct the latter.

But other philosophers reached opposite conclusions. Derek Turner (2005; 2007) claimed that historical hypotheses are usually underdetermined, i.e. the available evidence is insufficient to determine which of two or more competing explanations is true. Geological processes are typically information-destroying, erasing and deteriorating traces of past events. This process of epistemic wear makes the evidence we have access to insufficient to decide between competing hypotheses. Accordingly, local underdetermination problems are more widespread in the historical sciences, compared to the experimental ones. This remark clearly underwrites a different epistemic stance towards historical sciences, pessimism[7].

The debate has continued, with a major contribution from Adrian Currie's *Rock, Bone and Ruin. An Optimist's Guide to the Historical Sciences* (2018). As it is easy to infer from the subtitle, Currie explicitly frames his work in terms of opposition between pessimism and optimism. Currie's conclusion is that we should embrace epistemic/predictive optimism with regard to historical sciences: we should be confident that attempts at reconstructing the past will often succeed. He stresses how sciences such as archaeology and paleontology, far from being in a position of inferiority, have at their disposal a large array of epistemic strategies, from exploiting causal dependencies in the past to analogical reasoning and simulations, which enable them to overcome the limitations associated with their subject of investigation.

But why address such a problem in the first place? As Currie rightly points out, adopting an optimistic stance rather than a pessimistic one is not only a matter of philosophical preference. Scientists' view of their own discipline impacts their practices. An optimist's attitude encourages the continuation of scientific research, even in cases that appear particularly challenging. Fruitful speculation, the use of models, analogies and simulation, the hunt for new lines of evidence and the incorporation of new techniques and methods are all practices that make sense to adopt if we think they will get us somewhere. On the other hand, if we think pursuing a certain research question is hopeless, it is most reasonable to redirect one's efforts toward a different one. Moreover, whether we are pessimists or optimists regarding a particular scientific enterprise matters also for issues related to funding and priorities. As Currie points out in the introduction of his book, there is an impressive number of unstudied fossils in the collections of museums and

[7] See Tucker (2011) for the idea that the extent of underdetermination and overdetermination in the historical sciences cannot be established through philosophical reflection, but must proceed from an empirical analysis of the functioning of the individual sciences.

other research institutions. How many resources (if any) should we allocate to studying those remains? If we think studying them would not be particularly fruitful, then we will direct our funds elsewhere, or to other disciplines.

All of this is also relevant to the case of paleoanthropology. If we agree with Wood and Smith that we are not in a position to provide answers to the origin of unique human traits such as bipedality, then we should restrict our research to other issues: which species were probably bipedal, in which environments bipedalism evolved, what changes in the musculoskeletal system the evolution of bipedalism involved, and so on. This is a legitimate response to the problem posed by *why* questions. But we should not take the issue of pessimism lightly, since the consequences of how we view a discipline and its questions are profound.

In recent years, some of these philosophical reflections have been applied more specifically to paleoanthropology and paleolithic archaeology, whose evidential and methodological challenges make it an exemplary case among the historical sciences, and there is now a small, but lively community of philosophers working on the epistemology of paleoanthropology and human evolution (Currie et al., 2024). The topics addressed by philosophers are numerous: the analysis of inferences in cognitive archaeology (Pain, 2021; Killin & Pain, 2023), how to recognize different species in the fossil record, with particular reference to *sapiens* and Neanderthals (Meneganzin & Bernardi, 2023; Meneganzin & Stringer, 2024), the opportunities offered by the integration of a developmental perspective within paleoanthropology (Lequin et al., 2025), the identification of biological traits in physical anthropology (Meneganzin et al., 2024), the evolution of aesthetic and musical abilities in *Homo* (Meneganzin & Killin, 2024; Killin, 2018), the evolution of the concept of "behavioral modernity" (Meneganzin & Currie, 2022), and many others. Taken together, these contributions show that paleoanthropology is increasingly becoming a focus of philosophical inquiry in its own right—an ideal testing ground for debates about evidence, explanation, and the epistemic status of the historical sciences.

This finally leads to the aim of the present thesis.

4. The scope and structure of this thesis

The aim of this thesis is to examine evidential reasoning in paleoanthropology and related fields, particularly paleolithic archaeology. By "evidential reasoning" here we mean a crucial aspect of science in general and historical sciences more specifically: how data are transformed into evidence for our scientific claims. More specifically, how are fragmentary and degraded traces of the past (fossils, tools, aDNA, etc.) converted into evidence for reconstructions of our evolution?

What are the limits of evidential reasoning in paleoanthropology? How should our standards adapt to different hominin species? How does this interface with other aspects of scientific practice, such as the use of null hypotheses? How can the results of our reconstructions be integrated with other disciplines? These are some of the issues that we will try to address.

All this will be done, however, always keeping in mind two broader goals, one philosophical and the other scientific. From a philosophical point of view, the goal is to provide a contribution to the epistemology of paleoanthropology and human evolutionary sciences more generally. As we said, few philosophers have delved into the epistemology of paleoanthropology (yet) and there are still many methodological aspects to address. This is significant because paleoanthropology represents an especially compelling target for philosophical reflection. Like other historical sciences, it faces what Currie (2018) calls "epistemically unlucky circumstances" or conditions of "epistemic scarcity" (2021), given that its object of inquiry is in the deep past, and relevant evidence is sparse, degraded, and unevenly preserved. At the same time, however, it is also a "superstar science", as an archaeologist once described it to me during a conversation at the Max Planck Institute for Geoanthropology in Jena (see also Havstad, 2022, who characterizes paleoanthropology as a form of "sensational science"). Paleoanthropology, and the human evolutionary sciences more generally, attract considerable public attention, as they deal with the origins of a particularly self-interested species: us.

Regarding the scientific aim[8] of the present work, we have amply shown the variety of pessimist stances taken toward paleoanthropology. The goal is also to provide some guidance on how to answer these types of concerns. Taking a closer look at evidential reasoning in paleoanthropology should give us a better appreciation of the potentialities and limits of this discipline.

The methods adopted in this thesis align with the principles of contemporary naturalistic philosophy, which rejects any sharp division between science and philosophy. According to this view, philosophical inquiry should start from the findings and practices of the sciences rather than privileging *a priori* speculation over empirical research. Accordingly, each chapter will take as its starting point a specific case study and situate it within ongoing debates in paleoanthropology. When analyzing evidential reasoning, however, reference will also be made to the relevant philosophical literature.

[8] Which is, however, also philosophical. As Currie remarked, the question of pessimism and optimist attitudes toward science is a philosophically relevant one.

The structure of the thesis will be as follows. The first chapter examines how evidential standards are established and applied in paleoanthropology, using the recent controversy over deliberate burials in *Homo naledi* as a case study. It reconstructs how archaeologists have historically identified burials, from early archaeothanatology to the recent debate on *H. naledi*. Then the analysis of Berger and colleagues' preprints and their public peer review (2023) shows that disagreement does not concern which standards are relevant, but how they are interpreted. The chapter distinguishes between a data-checklist view, where standards mainly dictate what evidence must be gathered, and a scaffolding view, where the priority is given to the soundness of the inferential links between the data and the claim. The *naledi* team largely adopted the former, while critics enforced the latter, revealing a failure of adherence to community norms. Finally, the chapter extends this analysis to scientific communication. Overall, the chapter situates evidential standards as evolving, community-negotiated norms linking data, inferences, claims and even communication, setting the conceptual groundwork for the thesis as a whole.

The second chapter addresses the epistemological challenge posed by the remarkable variability of Paleolithic mortuary behavior and how evidential reasoning deals with this phenomenon. This variability gives rise to the *variability-of-expression challenge*: the challenge of reliably recognizing the same underlying cultural behavior when its material manifestations differ across time, space, and species. After providing an overview of variability in paleolithic burial practices, the chapter examines how paleoanthropologists and archaeologists define and identify burials. It then distinguishes several methods of inquiry for the individuation of burials: diagnostic lists of criteria, minimal definitions, null hypotheses, comparative and explanatory approaches, and analyzes how each responds to the problem of variability. In conclusion, the chapter argues that our epistemic strategies should factor in the variability of burial practices. Recognizing this implies adopting plural, adaptable strategies of inference, ones that preserve rigor while acknowledging diversity in how hominins expressed mortuary behavior.

The third chapter examines the use and misuse of null hypotheses in cognitive archaeology. Two recent proposals serve as case studies: Stibbard-Hawkes' (2025) "cognitively modern null", and Tennie et al.'s (2017) "no high-fidelity transmission null". Drawing on philosophical discussions of null hypothesis significance testing and "pseudo-nulls" (Bausman & Halina, 2018), the chapter argues that these proposals import statistical reasoning into a non-experimental domain where its application is not justified. While Tennie's null functions as pseudo-null, Stibbard-Hawkes' null functions in a way that is closer to a proper null hypothesis. Despite this, it still suffers from a serious problem that makes its application unprofitable: non-testability. The chapter proposes an alternative framework grounded in inference to the best explanation (IBE), a form of

reasoning already central to archaeological practice. The chapter concludes that null hypotheses, while legitimate in quantitative and experimental contexts, become problematic when transposed to historical sciences such as cognitive archaeology, where experimental testing is not possible and claims are supported by convergent long chains of inferences.

The fourth and final chapter asks how findings from paleoanthropology and archaeology should constrain neighboring disciplines that study human behavioral evolution, with a focus on evolutionary psychology (EP). After sketching EP's core claims and methodologies, the chapter argues that EP often under-integrates historical evidence when generating hypotheses about evolved cognitive adaptations. A central case study is the savanna hypothesis, according to which we should have an evolved aesthetic preference for savanna-like landscapes. Paleoecological and paleoanthropological evidence instead points to mosaic and highly variable environments, suggesting that adaptability and niche flexibility are more plausible targets than specialization for a single landscape type. The chapter concludes by arguing that evolutionary psychology's top-down and bottom-up heuristics should be more tightly constrained by archaeological and paleoanthropological evidence, raising the evidential bar and reducing just so storytelling, while shifting emphasis toward forms of context sensitive cognitive flexibility[9].

ACKNOWLEDGEMENTS

Doing a PhD is less of a solitary experience than one might think. Listing all the people who contributed, both outside and inside academia, is an almost impossible task, but I will do my best. First of all, I would like to express my heartfelt thanks to my supervisor, Vincenzo Fano, who has always been supportive and shown great enthusiasm for my project from the very beginning. Special thanks also go to everyone in the Synergia group in Urbino, with whom I have shared moments that have been deeply formative for me over the years. In particular, I would like to thank Pierluigi Graziani for his constant support and help and his willingness to guide me through the maze of the academic world. Niccolò Covoni, for the time we spent together, his friendship, and

[9] Almost all chapters feature final sections of "supplementary information" where it is possible to find useful resources that are not necessary for understanding the main arguments. In the case of the first chapter, there is an in-depth analysis of the claims made by the *naledi* team and the reviewers. In the third chapter, there is a series of paragraphs on cognitive archaeology, behavioral modernity, and parallels between comparative psychology and paleoanthropology. At the end of the fourth chapter, there is a brief discussion of the techniques that archaeologists and paleoanthropologists use to reconstruct paleoenvironments. I would like to thank the researchers of the Department of Archaeology at the Max Planck Institute for Geoanthropology for guiding me on this last topic.

his constant phone calls. Valerio di Lauro, again for his friendship, the work we did together, and the many conversations we had on astrobiology, contingency, and human uniqueness. And then Matteo Antonelli, Davide Pietrini, Ciro Saltarelli, Giovanni Macchia, Giovanni Galli, Simone Salzano, Flavia Marcacci, Adriano Angelucci, Stefano Calboli, Mirko Tagliaferri, Monica Tombolato, Gandolfo Vergottini, Vincenzo Nespeca, Ludovico Fusco and all the others who enriched these experiences of exchange and growth. I would also like to express my gratitude to the people who welcomed me with great generosity during my periods of research abroad, during a particularly itinerant year: Colin Allen, Andra Meneganzin, and Rocco Gaudenzi, who made every stage of this journey an opportunity for personal and professional growth. Especially to Andra, who had to put up with my doubts, my complaints, and numerous drafts of dubious quality. I would also like to thank all the people with whom I exchanged even just a brief chat during conferences, workshops, summer schools, and seminars. It may sound like a cliché, but the people I have met over the last three years will be one of my fondest memories of this PhD.

Finally, my warmest thoughts go to those who, although not part of my academic world, have contributed in an essential way to making the writing of this book possible with their simple but fundamental daily support. I sincerely thank my family, my friends, and, in a truly heartfelt way, Erika.

To all of you, truly: thank you.

CHAPTER I
BURIED, OR MAYBE NOT: EVIDENTIAL STANDARDS ADHERENCE IN PALEOANTHROPOLOGY

1. Introduction[10]

We will begin our analysis of evidential reasoning in paleoanthropology with a case that will recur throughout the following chapters: the debate over the alleged burials of *Homo naledi*. This small-brained hominin, first discovered in 2013, became the center of renewed controversy in 2023, when the research team released a series of preprints presenting evidence for intentional burials and rock engravings. What makes this case particularly significant, and in part inspired the very premise of this thesis, is that the ensuing debate unfolded publicly. The journal chosen by the *naledi* team, *eLife*, recently adopted an open publication model in which preprints are published alongside peer reviews and the authors' reply. This format exposes dimensions of reasoning in paleoanthropology that would otherwise remain largely inaccessible. In particular, this case brings to light epistemologically significant aspects of scientific practice, such as meta-reflections on evidential standards, as well as social dynamics within the community, including the enforcement of norms on those perceived as "rebels". Moreover, many scholars joined the discussion, often critically, through journal articles, commentaries, videos, podcasts, and other media outlets. The *Homo naledi* controversy therefore provides an exceptionally rich set of materials, offering philosophers a rare chance to observe, in real time, how both the epistemic and social dimensions of paleoanthropological practice unfold and interact.

Finally, this debate concerns an area that has not yet been explored by philosophers: burials. Although recent years have seen increasing interest in how animals perceive death, both in animal cognition and philosophy (see for example Gonçalves & Carvalho, 2019; Monsó, 2024), little philosophical attention has been devoted to the evolution of mortuary practices in our own species (but see Sheets-Johnstone, 1986, for an exception). One of the ways humans have sought to cope with the drama of death is through burial, a practice now taken for granted across numerous civilizations. Yet important questions remain open: when did burial first emerge, in which species,

[10] This chapter stems from a collaborative work with Andra Meneganzin. A manuscript based on it is currently under review at *Biology and Philosophy*. I would like also to thank the Ramsey Lab, and especially Grant Ramsey, Alejandro Fábregas Tejeda, Gianmaria Dani and Juncheng Tao, for providing valuable feedback and suggestions on the manuscript in preparation.

and under what pressures, selective or otherwise? These are precisely the questions that funerary archaeology seeks to address.

The study of Paleolithic burials is also methodologically challenging. Corpses were often placed in shallow pits dug with rudimentary tools, or even in natural depressions rather than purposefully prepared graves. Over millennia, taphonomic processes have further obscured these contexts. The central task, therefore, is to determine whether a body was deliberately interred or whether it simply remained where death occurred (natural deposition)[11]. The study of burials is therefore a paradigmatic case of evidential reasoning, and of how fragmentary and degraded data can, with great epistemic effort, be transformed into evidence for statements about the behavior of our ancestors.

This chapter is structured as follows. Section 2 provides an overview of the work done by philosophers on standards of evidence and similar concepts. Section 3 offers an overview of how burials are identified in the archaeological record and how standards for assessing evidence have emerged in archaeological practice. Section 4 lays out the claims of deliberate burials in the *Homo naledi* case. Section 5 articulates what underwrites the lamented failure to meet the community's epistemic standards by drawing from response articles and public peer review—what we'll call "adherence failure". Specifically, we will show that adherence failure is chiefly driven by the contextual application and interpretation of criteria and standards rather than by a failure to identify them. Next, Section 6 will discuss how adherence failure also applies to best practices in communicating knowledge claims both within paleoanthropology and to the broader public. Section 7 briefly considers the risks that come with adherence failures. Section 8 concludes.

2. Evidential standards, a concept yet to be explored

Philosophers have long departed from conceptions of science as axiomatic, unified, and stable systems of inquiry. Underlying the production of scientific knowledge are highly dynamic, local, and historically situated collective processes. Among the factors shaping the dynamics of knowledge generation, epistemic standards play a crucial role (Brigandt, 2012; Jacoby, 2023). Epistemic standards inform methodological choices (Cartwright, 2007), the configuration of data as evidence for scientific claims (Leonelli 2015, 2019), inferential strategies (Chapman & Wylie, 2016), and the formation and evolution of criteria of explanatory adequacy (Love, 2008). In

[11] From this perspective, identifying burials from more recent periods is generally easier. The construction of special structures to house the body, as well as a certain systematization of, for example, grave goods, makes the context less ambiguous. For example, the Neolithic site of Barma Cotze, which I visited in August 2025, featured a series of stone cists that we know were used as tombs by Neolithic shepherds in the Aosta Valley, so their recognition as burials is fairly straightforward.

contexts of inquiry marked by "epistemic scarcity" (Currie, 2021)—e.g., in sciences like paleontology, archaeology, or paleoanthropology which have impoverished access to their objects of study—setting plausible standards would seem key to squeezing information out of sparse remains and securing epistemic success in the face of the eroding action of deep time.

Debates in paleoanthropology and human evolutionary studies frequently evoke "evidential standards" that would be required to support knowledge claims. That is, paleoanthropologists' and archaeologists' prospect of reconstructing long-gone lifeways and capacities of extinct minds appears to critically turn on the identification and adequacy of relevant evidence, and how this is mobilized to make old bones, remnants, and degraded DNA tell their stories. However, it remains unclear in such epistemic contexts what evidential standards ultimately amount to and, crucially, how to assess failure in meeting them. While philosophers have attended to the implications of notions of (or akin to) "evidential standard" in other scientific contexts (notably, in the biomedical sciences, see Stegenga, 2011; Tempini & Leonelli, 2021), less attention has been devoted to its conceptual nature, or its role in sciences like paleoanthropology and archaeology (but see Chapman & Wylie, 2016; Currie et al., 2024, Meneganzin & Killin, 2024 for broader discussions).

Here we leverage recent discussions on claims of deliberate burial practices among small-brained hominin *Homo naledi* (Berger et al., 2023a) to make some headway into how evidential standards are constructed, interpreted, and enforced by practicing paleoanthropological communities. As such, our focus will be on claims that infer behavioral and cognitive capacities from aspects of material evidence—mainly anatomical and geological. The stakes in the *Homo naledi* case are non-trivial. Evidence that a small-brained species engaged in funerary behaviors and meaning-making activities would generate a cascade of significant implications, from retrodating the evolution of such behaviors to reassessing the causal weight of encephalization in hominin evolution (Fuentes et al., 2023). As we will show, the debate surrounding the capacities of *Homo naledi*, along with the specific publication dynamics underlying key papers, provide a paradigmatic case study in which paleoanthropological reasoning and standards for licensing core inferences explicitly come to the forefront. This presents a *sui generis* opportunity for philosophical analysis.

3. Recognizing burials in the archaeological record

Among the many marks of cognitively and socially rich lifeways, mourning behavior is an oft-invoked feature. Behavioral expressions and practices of engaging with the loss of a close group member offer deep insights into the emotional, psychological and cultural lives of agents.

Mourning, however, is not peculiar to the genus *Homo* alone. Animals like chimpanzees and elephants are not limited to immediate responses when confronted by the death of a group member —"grieving". They also engage in time-extended processes aimed at maintaining connections with the dead, such as prolonged dead-infant carrying or death sites visiting—"mourning", properly speaking (Marrone, 1998; Stiner 2017).

In paleoanthropology and Mid-Upper Paleolithic archaeology specifically, the identification of burials as a distinctively human mourning activity still plays an important role[12]. Burials are just one among many mortuary practices, which show considerable variation both in the present and in the deep past. As far as archaeological discussion is concerned, burials are a time-consuming and labor-intensive option for dealing with the dead, especially in light of more inexpensive alternatives such as simple abandonment, setting corpses adrift in rivers, or caching them in natural fissures. Crucially, while the emotional components and intellectual beliefs that underwrite mourning and mortuary practices may not be archaeologically visible, burials are. This notwithstanding, their recognition is far from straightforward. While burial practices among late Neanderthals and *Homo sapiens* from around 120-100 ka are established on increasingly solid footing (Stiner 2017, but see Sandgathe et al., 2011; Dibble et al., 2015 for contextual criticism), claims of mortuary behavior for earlier periods are known to be more difficult and to lack comparable credibility, especially at sites excavated before the advent of current excavation and recording methods (Pettitt, 2018). So, how are burials recognized in practice?

Over the past decades, paleoanthropologists and archaeologists have confronted several evidentiary contexts in which burial features have been established based on discussions of explicit criteria. One context where burials are still scarce and lack well-dated evidence is the Middle Stone Age (MSA) in Africa. Recent work by Martinón-Torres and colleagues (2021) has claimed to have described the primary burial of a 2.5–3-year-old child recovered from the MSA layer of the site Panga Ya Saidi (PYS) in Kenya, and dated around 78.3 ± 4.1 ka. Burials are usually defined as the intentional excavation of a pit, the placement of a body inside of it and the infilling of the pit with soil to protect the body (Pettitt, 2011; Martinón-Torres et al., 2023). A "primary" burial is defined as the disposal of a fresh body in its final resting spot, where the decomposition process takes place. This is usually contrasted with "secondary" burials, where dry bones are transported to their final resting spot (Duday, 2006). To support the claim that the child ("Mtoto") was indeed buried on the spot, the team provided the following justification (the passage is worth quoting at length):

[12] But see Pettitt (2018) for arguments in favor of a nuanced understanding of early human morbidity beyond the search for burial evidence alone.

> Four features demonstrate that a fresh body was placed in a location where the entire process of decomposition took place: (a) the macroscopic anatomical integrity of the body, especially unstable articulations; (b) the minimal displacement of bones, with movements explained as a consequence of decomposition; (c) the abundance of terrestrial gastropods that feed on earthworms in close proximity to the corpse; and (d) geochemical and histological analyses that indicate in situ decomposition and putrefaction processes. The PYS child remains meet all four of these criteria (Martinón-Torres et al., 2021, pp. 96-97)

In listing these criteria, Martinón-Torres and colleagues refer to the seminal work of archaeothanatologist Henri Duday (2006), who presented a comprehensive set of standards or "characteristics" required to identify primary burials. These include a) anatomical articulation, in particular of unstable joints; b) decomposition fluids and the decomposition of the burial environment; c) the displacement of bones within the cadaver in the course of decomposition; d) the original position of the body. Martinón-Torres and colleagues' work instantiates these criteria most glaringly.

Therefore, to establish that some human remains have been intentionally buried, several factors must be on the radar. The field archaeologist must engage in detailed documentation of single bone elements or fragments *in situ*, their exact position, and orientation within the body's anatomical architecture. As regards osteological data, joints prove to be particularly informative. If a body is placed in a grave with all the organismal structures maintaining articulations intact (from skin to tendons and ligaments), contiguous bones might be prevented from disjoining. The integrity of fragile articulations that are much more prone to displacement (involving bones of small size or weak ones) is seen as an important indicator of the integrity of the overall cadaver. But the bones must also be documented within the context of the grave's architecture and environment. Among the many taphonomic agents acting on the disposition and coherence of bones, the law of gravity stands as the most obvious. The movements generated by the decay of soft tissues vary based on the original disposition of the body, with head rotation impacting the weight of the cranium during the deterioration of cranio-vertebral support. There are oft-encountered modifications, such as the flattening of the rib cage, the dislocation of the vertebral column, or the collapse of the pelvic girdle (Duday 2006). Further disturbances are generated by the infiltration of sediments with putrefaction fluids, the action of earthworms and insects concentrating around the corpse, and their respective predators (insectivores, amphibians, and reptiles), often leaving traces in the burial area. Mtoto and the good preservation of environmental proxies at PYS are informationally rich along many of the abovementioned axes.

As such, paleoanthropologists and archaeologists, when faced with complex and ambiguous trace data, develop explicit heuristic strategies for inferring highly valued behaviors such as

burials. As scientific practice is a dynamic and evolving enterprise, it should come as no surprise that these strategies become more stringent and explicit over time and are integrated with additional methodological resources. The PYS team, building on established archaeothanatological practice and Duday's criteria, also applied geochemical and histological analyses to assess the *in situ* decomposition of the body and investigated the presence of a pit through detailed granulometric and stratigraphic tests. Consequently, complex archaeological reconstructions as burial identification come to gradually benefit from the example of well-established cases, providing newly built standards against which novel burials can be assessed. As we'll see, understanding what these standards are and how they work is critical in appraising this process. A case where novel burial claims have been attempted is with *Homo naledi* and the Dinaledi Chamber, to which we'll turn now.

4. *Homo naledi*: claims for mortuary behavior

In 2013, a team led by paleoanthropologist Lee Berger discovered the skeletal remains of at least 15 individuals in the Dinaledi Chamber of the Rising Star cave system in South Africa. The remains belonged to a previously unknown species, whose in-depth description and taxonomic baptism was first provided a couple of years later: *Homo naledi* (Berger et al. 2015). The morphology of *Homo naledi* is remarkable for its original mixture of primitive, australopithecine-like and derived characters. The lower limb, particularly the ankle and the foot, exhibits a largely human-like morphology, whereas the pelvis is flared, resembling that of australopiths. The hands and wrists are broadly comparable to those of other *Homo* species, although the fingers are markedly curved, as in *Australopithecus*, a feature plausibly associated with climbing and a degree of arboreal locomotion. The morphology of the shoulder girdle and the ribcage likewise recalls australopith anatomy. Cranio-dental morphology displays a similar mosaic pattern. The estimated weight and height are comparable to those of modern small-bodied human populations. The cranial capacity (between 465 cc and 560 cc) falls completely within the range of variability of australopiths, although the cranium is structurally similar to those of early species of *Homo*. Taken together, these features suggest that *Homo naledi* was adapted to a distinctive niche. The species appears to have combined adaptations for climbing with evidence of efficient and derived bipedal locomotion; its hands and wrists were capable of precision manipulation while retaining traits associated with arboreal activity; and, despite its small brain size, its cranial organization exhibits features characteristic of *Homo*.

The discovery of *Homo naledi* is even more surprising when its chronological placement is taken into account. Geochronological analyses have shown that the fossils are unexpectedly young, dating to between approximately 335,000 and 236,000 years ago (Dirks et al., 2017). This age range is considerably more recent than many researchers initially anticipated, especially given *H. naledi's* anatomically primitive traits. The revised chronology places *H. naledi* in the late Middle Pleistocene, indicating that these small-brained hominins lived in southern Africa at roughly the same time as early *Homo sapiens* and other large-brained hominin species. This finding substantially expands the known diversity of the genus *Homo* during this period, showing that a lineage retaining markedly archaic features persisted until relatively late in prehistory alongside more derived human forms.

Other elements of the discovery are noteworthy. The Dinaledi chamber is a deep and isolated cavity located approximately 30 meters below the surface and more than 80 meters from the nearest present-day entrance. Access to the chamber requires navigating a complex network of narrow passages, including a vertical chute measuring in places less than 25 cm in width, and involves complete darkness throughout. In addition, the assemblage included no stone tools, no cut-marked animal bones, and only a few isolated bird and rodent bones that could not be linked to the hominin fossils. The unusual context has played a central role in subsequent debates concerning the processes responsible for the accumulation of the bodies.

Initially, the team proposed that the accumulation of bodies within the complex Rising Star cave system was probably due to deliberate disposal acts by *Homo naledi* (Dirks et al. 2015). The bodies would have been either dropped or carried into the chamber, though scenarios involving mass fatality or a death trap were also considered alternatives worthy of investigation. On the other hand, the geological and taphonomic context appeared inconsistent with common mechanisms known from other South African cave sites, such as carnivore accumulation, fluvial transport, or occupation by hominins. The chamber is located deep within the cave system, permanently dark, and inaccessible to large non-hominin animals, while the assemblage consists almost exclusively of *Homo naledi* remains, with no evidence of predation, scavenging, or transport-induced abrasion (but see Val, 2016). The skeletal material shows limited weathering and suggests that individuals entered the chamber largely intact, with disarticulation occurring in situ during decomposition. Subsequently, the team partially revised their interpretation.

In July 2023, the team published three preprints in the scientific journal *eLife.* The preprints followed the journal's new publication process, which requires them to be published along with public responses from reviewers, the journal's assessment and, eventually, the authors' response

(*eLife,* 2022)[13]. The public unfolding of peer review is crucial to the present work, as it allows us to analyze relevant aspects of the scientific debate that would have otherwise remained inaccessible. It is on the publicly available materials that this chapter focuses, with no suggestion that these exhaust the complex trajectory through which scientific claims take shape and come online[14]. Further, as the debate is still unfolding, we focus here on its early stages, while also touching on more recent developments later in the chapter. This focus is also theoretically motivated: closely examining the initial presentation of, and responses to, a hypothesis aids in appreciating its epistemic development, thereby helping to distinguish genuine changes from points of continuity.

Of the three 2023 preprints, the first—on which this chapter focuses—argues that at least some of the bodies were intentionally buried in pits dug into the cave soil (primary burials), rather than simply abandoned (Berger et al., 2023a). This would imply the adoption of complex behavior, tied to discussions on symbolism and other "behavioral modernity" markers (see Scerri & Will, 2023 for a recent overview), by a hominin with a cranial capacity comparable to that of an australopithecine. Further, this would constitute the earliest known evidence of a burial ever recorded. How is this conclusion reached? In what follows, we'll first offer a brief overview of these initial claims, before turning to an analysis of the debate these have stirred.

The preprint presents the geochemical and osteo-anatomical analyses carried out on three "features" that the authors suggest indicate primary burials. These include two *partially excavated* features in the floor of Dinaledi Chamber (Dinaledi Features 1 and 2) and another feature from the floor of Hill Antechamber, which was *extracted en bloc* and encased in a plaster jacket, but not excavated (Hill Antechamber Feature).

Crucially, the team discusses "five key observations" that would support the interpretation of the three features as instances of primary burials:

1. the difference in sediment composition within the feature compared to surrounding sediment*;*

[13] It is worth noting that the adoption of the new publication process prompted Clarivate, the analytics company that manages the Web of Science database, to remove the journal's impact factor: https://www.science.org/content/article/web-science-index-plans-end-elife-s-journal-impact-factor

[14] In response to a remark made by one of the authors of the preprint (private communication), we acknowledge that there may be a complex background of conversations and decisions that occurred prior to the release of the preprint, which are not reflected in it. However, we believe this is true for most scientific publications. There is often a long trajectory of input—such as conference presentations, informal exchanges, private feedback, participation in peer review of colleagues' work—and even when this is incorporated into the published version, the work may still stir controversy. We think that it would be problematic to discount the value of the published work on these grounds.

2. the disruption of stratigraphy;
3. the anatomical coherence of the skeletal remains;
4. the matrix-supported position of some skeletal elements;
5. the compatibility of non-articulated material with decomposition and subsequent collapse (Berger et al, 2023, p. 6).

Let us briefly see what evidence is adduced to support the five observations[15]. Evidence for the first observation consists of X-ray diffraction (XRD) and fluorescence (XRF) analyses comparing sediments within Dinaledi Feature 1 to those from the surrounding chamber, subsequently subjected to principal component analysis (PCA)[16] to capture the variation within the datasets. The second point is supported by the disruption of a laminated orange-red mud layer visible in profile near Feature 1 and in CT scans from the Hill Antechamber. For the third observation, the authors argue that the spatial arrangement of skeletal remains in Feature 1 indicates a mostly intact single individual. For the Hill Antechamber Feature, again, the main source of evidence are the CT scans, as the feature remains unexcavated. The fourth point relies on the vertical or sub-vertical positioning of bones, such as cranial and rib fragments, interpreted as matrix-supported, both in Dinaledi Feature 1 and Hill Antechamber Feature. Finally, for the fifth observation, the authors claim that the displacement of some non-articulated bones is compatible with post-depositional collapse following soft tissue decomposition.

It is immediately clear that the paper's central claim vitally depends on what to make exactly of the above-listed elements—their role, identification and specific interpretation, as we shall discuss. After presenting the data supporting the five observations, the concept of evidentiary standards explicitly comes up, with the authors stating that, in light of their findings, the «burial features meet evidentiary standards used for recognizing burials of *H. sapiens*» (p. 17). The interpretation of the "five key observations" as addressing evidential standards is further echoed in the responses to the reviewers: «[w]e feel we have met the standards of demonstrating burials in *H. naledi*» (p. 85). In this regard, and related to their identification, the authors cite Mtoto's burial at Panga Ya Saidi (see our Section 2; Martinón-Torres et al., 2021), implying its acknowledgement as providing a reference point—a "gold standard", as it were—for the recognition of primary burials.

[15] Besides the key points of contention that we'll touch upon in the following section, we refer the reader to the original paper and the reviewers' responses for broader context, which cannot be fully recapitulated here.

[16] A principal component analysis (PCA) is a common analytical method used to reduce the dimensionality of a large dataset while retaining most of its original information.

Prima facie, we can see how the standards identified by Berger and colleagues appear to bear resemblance to those of Martinón-Torres and colleagues (which in turn build on Duday 2006). This is not to say that there are no differences. For instance, Berger and colleagues make no mention of remains of gastropods or other animals feeding on earthworms. Finding the remains of these critters is an important indicator that the body decomposed at the location where it was found. Assessing the presence of surface modifications on *H. naledi* bones consistent with the activity of these terrestrial gastropods has proven problematic in the past, as this datum is leveraged by Val (2016) to argue that the bodies likely decomposed on the surface or in other parts of the cave. To date, there is still no evidence of snail shells associated with *H. naledi* remains. However, except for this point, there seems to be a convergence on the standards for assessing burial behavior first identified by Duday (2006). Thus, if evidential standards are key to presenting evidence of deliberate burials in *H. naledi*, they equally are to those unconvinced by the paper's conclusions. In the initial stage of the debate, strong skepticism was raised towards the conclusions reached by the team. This is reflected by assessments from the openly accessible referee reports: «In its current form the paper (...) does not meet the standards of our field» (Reviewer #1, p. 70), «I feel that there is a significant amount of missing information in the study presented here» (Reviewer #2, p. 72) «important lines of evidence are lacking» (Reviewer #3, p 73).

A question then arises: Where exactly does this disagreement stem from? To answer it, we will analyze and provide a general articulation of the reviewers' lines of criticism opposing the case for burial behavior.

5. Interpreting evidential standards

One may think paleoanthropology is so rife with controversy and disagreement because it lacks shared standards to which scientists can appeal to guide evidence collection and hypothesis evaluation, or because it misaligns with the practice of disciplines where those standards are established (see Tattersall, 1992 for discussion on taxonomic practice). While the scarcity of trace data is a common trope, vigorous expressions of pessimism toward paleoanthropology, both from insiders and outsiders of this discipline, appear to stem, at least in part, from how we handle the data that we do have (e.g., Smith, 2016). Thus, we might think that in cases where explicit criteria are formulated, disagreement should be either significantly reduced or, when present, explained via a lack of convergence on the normative power of those standards. In what follows, we will show that this reading proves unhelpful in making sense of the *H. naledi* debate. To do so, we will take a closer look at what grounds skepticism regarding the claim that *H. naledi* buried its dead.

5.1 Reviewers' assessment

In the debate surrounding *H. naledi'*s purported burials, the analysis and the conclusion reached by the team have been subject of much criticism (see the reviews on Berger et al., 2023a; Fuentes et al., 2023; Martinón-Torres et al., 2023; Foecke et al., 2024; Pettitt & Wood 2024; and, retrospectively, Val, 2016; Stiner, 2017; Egeland et al., 2018). Surprisingly, however, none seem to target the evidential standards themselves as advanced by Berger and colleagues, but—as we shall see—different aspects of the investigation. To be fair, as noted in several places, the 2023 preprint included a relatively limited bibliography (32 references). However, the dispute does not appear to hinge on a *misidentification* of key archaeothanatological principles—for instance, there is no suggestion that the reference to Martinón-Torres and colleagues (2021) is misleading. While some critiques do point to a lack of engagement with the broader archaeothanatological literature, much of the discussion takes place against the backdrop of the minimally stated standards themselves (e.g., R4; Martínon-Torres et al., 2023, Foecke et al., 2024).

Although the lines of criticism are numerous and varied, it is possible to identify a few core categories. These include: (i) quantitative inadequacy of data; (ii) qualitative inadequacy of data; (iii) lack of primary analyses; (iv) lack of eliminative reasoning; (v) biased communication. Before addressing these points in detail, two brief clarifications are in order. First, when discussing data in the context of a historical science like archaeology, we'll refer to a body of collected information that can refer to either primary physical features of objects (the bones, the traces and the marks left by ancient humans; on fossil data see also Watkins, 2024) or non-physical epistemic objects (that is, "secondary data" like measurements and recordings of the former) (Leonelli, 2017). Second, while our categorization of the critiques is motivated by conceptual distinction and clarity in presentation, we acknowledge that these often become entangled with one another in the empirical debate. Let us turn to them now.

(i) *Quantitative inadequacy of data*. "Quantity" requires qualification. Our concern here is not merely a simple lack of data abundance, but aligns with what Sabina Leonelli and colleagues call "data shadows" (Leonelli et al., 2017): the absence, unavailability or inaccessibility of relevant data. As they note, such claims only make sense if accompanied «by a clear vision of what data would be desirable and why and by demands for additional activities, resources, or skills to bring such data into the light» (Ivi, p. 194). Indeed, several reviewers point out that key geochemical and anatomical data required to support the burial hypothesis are missing. This is often presented as a consequence of dealing with features that, as we saw, are not fully excavated, thus limiting the accessibility of desirable information (cf. R2, R3, Martinón-Torres et al., 2023). Recent work

by Foecke and colleagues offers an in-depth analysis of the geochemical and sedimentological data (Foecke et al., 2024) arguing that the data presented, together with the analyses conducted, are *insufficient* to support the *H. naledi* team's conclusion. In other words, key data desirable for supporting a burial claim remain in the shadows—raising the further question of why they are absent[17]. In Foecke and colleagues' words: «there is not enough information about the sediment or fossil material properties to determine if specific elements represent a fossiliferous signal exclusively» (p. 5). With fundamental anatomical and sedimentological information lacking, burial claims appear to stand on shaky ground.

(ii) *Qualitative inadequacy of data.* Some data may be missing, but what is available could still be informative and adequate for the research at hand. Initial criticism of the burial claims appears to address both of these aspects. We interpret these as claims regarding both the *resolution* and *adequacy-for-purpose* of data (Bokulich & Parker 2021). Concerning the former, claims regarding, for example, the disruption of stratigraphy ("observation 2"), do not seem to be supported by data with sufficient resolution—that is, they lack the necessary level of detail for reviewers and commentators to clearly identify or appreciate the reported disruption (see R4). This is further reflected by comments on the limited interpretability of figures meant to support such a claim (e.g., Figure 3B and the laminated orange-red mud clast, LORM; see also discussion in Section 5). Further, despite missing data (point i), there have been attempts to replicate the PCA results used to identify a difference in composition between the sediments contained in the feature versus those outside of it ("key observation" 1). On several accounts, the PCA results failed to replicate, suggesting that the underlying data lacks stable structure (cf. Foecke et al., 2024; R4 of Berger et al., Fuentes et al.).[18] "Adequacy-for-purpose" (*sensu* Bokulich & Parker 2021) indicates instead that some data are incapable of fulfilling the epistemic role required from them in this particular context. For instance, it has been noted that, to demonstrate that a pit was deliberately excavated, adequate data should show differences between the sediment inside the pit and that outside it. However, the data provided only track expected differences between fossil-bearing and non–fossil-bearing matrices, without identifying any infilled sediment. As such, they are deemed inadequate for demonstrating the existence of a pit (cf. Martinón-Torres et al., 2023, Section 2.1).

[17] This quote by Leonelli and colleagues is particularly telling: «It is of course possible to interpret the phenomenon of shadowing in a linear fashion, where light represents knowledge (or at least the means to it) and shadows thereby its absence. However, as exemplified by any museum exhibit or art gallery, shadows can be used to hide things as well as to make things more clearly noticeable or to emphasize aspects of an object on which one wants to draw attention.» (Leonelli et al. 2017, p. 194).

[18] Apparently the lack of replicability concerns non only the PCA «the XRF and PSD data as provided by Berger et al. either do not support their conclusions when analyzed using standard techniques or are not reported completely (impeding replication)» (Foecke et al 2024, p. 7).

(iii) *Lack of primary analyses*. The data that are "shadowed" clearly impact the availability of analyses that depend on them and, consequently, the information that can be extracted. Recall, once again, that the burials are not fully excavated (only Dinaledi 1 has been partially excavated). While the authors justify this omission by aiming to avoid unnecessary damage to the features and preserve them for future studies (see their response on the three 2023 papers), reviewers argue it undermines the possibility of, among other things, carefully studying the skeletal remains to ascertain their integrity, determine whether they belong to one or more individuals and assess their spatial placement relative to other bone elements. But core methods and analyses are missing even for those data that are supposed to be "under the light". The absence of micromorphological analysis—an essential technique for reconstructing the depositional context of a site—appears as a prominent example (R1, R2 and R4 of Berger et al., 2023a; Martinón-Torres et al., 2023, section 2.2; Foecke et al., 2024). This is especially conspicuous for the much-debated information meant to support the observation of a disrupted stratigraphy (that is, the profile depicted in Figure 3b and the LORM layer). Curiously, both techniques (micromorphology and complete excavation) are employed in Martinón-Torres and colleagues' paper on the Panga Ya Saidi child, the case study that the *H. naledi* team claims to take as a model for their evidential standards (see their response in 2023a, p. 82).

(iv) *Lack of eliminative reasoning.* There is the data that is made available, that which remains inaccessible, and the types of analyses that each of these conditions enable or preclude. A further, crucial issue concerns the inferential moves made by the *H. naledi* team on the basis of this data foundation. Several commentators have consistently highlighted a failure to consider and possibly eliminate alternative hypotheses that are equally compatible with the proposed data. This is especially apparent in instances where several lines of evidence (or, more accurately said at this stage, data) are subjected to the problem of equifinality: the same result can be explained by very different causes. The problem of "unconceived alternatives" in philosophy is a notorious one: the ghost of the alternative explanations that we have not yet imagined undermines the credibility and stability of our best hypotheses (Stanford, 2006). Curiously, in the present case, it is more the long-conceived alternatives that pose a challenge in the absence of adequate engagement with eliminative reasoning—the gradual elimination of alternative but compatible explanations to restrict the space to plausible, rival hypotheses (Forber, 2011). For example, a certain degree of skeletal cohesion may indicate that the bodies were somehow "protected". However, in an environment such as the Rising Star Cave System, where, according to the team itself, there are few natural agents capable of scattering skeletal remains, it is not surprising to find bodies that maintain a certain level of integrity. Therefore, mobilizing this data as evidence that the remains were deliberately buried is not sufficient; further work is needed to eliminate alternative scenarios.

While the team claimed to have ruled out some alternative hypotheses in the past, such as water transport, death trap and passive gravity-driven accumulation (see Dirks et al., 2015; Dirks et al., 2016; Hawks et al., 2017; Elliott et al., 2021; Brophy et al., 2021), some reconstructions still remain plausible. These include long-stated scenarios of natural accumulation of remains due to erosion, sediment slumping and drainage or the activity of carnivores (see especially R1 and R3 on Berger et al., 2023a; Martinón-Torres et al., 2023; and *ante litteram* Val, 2016; Egeland et al., 2018).

Finally, (v) biased communication is identifiable on multiple levels—from the visualization of results in the original paper, to the use of non-standard reporting methods, and a deliberately persuasive, biased style of argumentation. We will address these issues separately in Section 5 and show how they relate to differing interpretations of evidential standards.

Having briefly articulated some of the core issues surrounding the burial interpretation, in the following section, we will suggest an epistemic diagnosis for the disagreement between the *H. naledi* team and the critical paleoanthropological community.

5.2 Conceptions of "evidential standards"

Thus far, we have observed that claims in favor of *H. naledi* burying their dead present several critical issues, according to public peer review and other peer assessments. Skepticism following the publication of the 2023 preprint stems from shortcomings in the data and arguments cited in support of intentional deposition: important data remain "shadowed" (absent, unavailable, or inaccessible); available data lack sufficient resolution or are unfit-for-purpose; eliminative reasoning is poorly employed even in the presence of long-introduced hypotheses; and finally, communication is biased. Recall that the 2023 burial preprint opens the discussion with the following statement: «These burial features meet evidentiary standards used for recognizing burials of *H. sapiens* (Martinón-Torres et al., 2021)». This leaves us with a puzzle: if the issue is not with the standards themselves, then where does the disagreement lie? In this section, we argue that the divergence stems from how these standards are conceptualized, and, consequently, from differing implications on what is required for them to be met. In light of the views of *H. naledi* team and the reviewers' comments, let us now outline two different conceptions of "evidential standards." We'll call the former the "*data checklist*" approach and the latter "the *scaffolding* approach". We argue that these two different construals respectively track the attempted evidential reasoning in the *Homo naledi* papers and the reasoning invoked in the public peer-review and

critical papers. The interesting upshot, as we'll see, is that the burial claims ultimately fail on both accounts.

Before doing so, it is necessary to refresh the conceptual difference between data, evidential claims, and what licenses and delimits inferences from the former to the latter—particularly warrants and rebuttals. This distinction draws primarily from influential philosophical work on evidential reasoning in the historical sciences (Chapman & Wylie, 2016, drawing on Toulmin's famous 1958 schema; Kokkov, 2019), but aligns also with aspects of the pragmatist theory of evidence (Reiss 2015), as well as with philosophical accounts of data and their treatment as potential evidence (Leonelli, 2015, 2016; Bokulich & Parker 2021)[19]. As stated earlier, we used "data" to refer to both primary physical features of objects and non-physical epistemic objects. Most philosophical discussions converge in the treatment of data as "potential evidence" (Leonelli 2015), or as being ascribed "evidential value" (Leonelli, 2016), being subject to different uses as evidence (Bokulich & Parker, 2021), or providing support for an evidential claim, that is, a conclusion being defended (Chapman & Wylie, 2016; Toulmin, 1958). All of this occurs by means of other fundamental epistemic activities and items. In some influential articulations, these include the warrants, which are fundamentally linked to the inferential aspects of argumentation— that is, how and/or why the data support the conclusion — and variously draw from middle-range theories, background and domain-specific knowledge (cf. Toulmin, 1958; Chapman & Wylie, 2016; Reiss, 2015; for discussion of middle range theories, see Binford, 1977; Jeffares, 2008; Currie, 2018). The rebuttals, in turn, indicate at which steps the argument is defeasible and where the warrants may be subjected to counterarguments, exceptions and alternative scenarios compatible with the data. Chapman and Wylie notoriously referred more generally to "inferential scaffolding" to indicate what is required to mediate data as evidence, and included among these gap-crossing assumptions, auxiliary hypotheses, and background knowledge (Chapman & Wylie, 2016, p. 35). It is on the construction of these scaffolds that archaeological reasoning ultimately depends. Far from being inescapable or determined by logical necessity, evidential claims are assessed on the reliability of these scaffolds, their ongoing scrutiny as field knowledge grows, and their fitness for purpose. Let's turn now to our distinction between two possible conceptions of evidential standards.

The "data checklist" approach interprets evidential criteria primarily as a list (or set) of data points to be gathered in support of a particular claim. In this sense, evidential standards would provide normative guidance on what data should be collected. Under this view, the criteria discussed in the Panga Ya Saidi case and the broader archaeothanatological literature would supply

[19] While these accounts differ in important aspects, those differences are not central to the present discussion.

guidelines for recovering and recording *type* traces: skeletal elements, geological and stratigraphic data, etc. What is immediately evident from this approach is that aspects related to the warrants of reasoning—the claims that license the inference from data to the conclusive evidentiary claim — risk being flattened onto those of support—the collection of sufficient and relevant type data. Adequately supporting a claim could then be easily conflated with providing valid warranting arguments. Further, and crucially for our purposes, on the checklist conception, disagreement should be absent (or nearly so) whenever explicit evidential criteria are involved. Given their nature, and if they are shared by the relevant community, it should be relatively unproblematic to determine whether a claim is warranted: one need only check whether all the relevant data types dictated by the criteria have been collected. Moreover, if meeting primary data requirements is what matters chiefly, these criteria should also generalize across various epistemic, temporal, and spatial contexts.

Put this way, this conceptualization is philosophically dire. As such, it may reflect folk epistemological, rather than philosophically informed notions of how evidential criteria are supposed to work. However, a more charitable interpretation is also possible: that it reflects a prioritization of aspects of the investigation over others. Giving priority to data collection can be particularly pressing in the historical sciences, where the recovery of traces from the past is both essential and extraordinarily resource-consuming. Especially when attempting to infer highly valued behavioral and cognitive capacities of long-lost hominin forms, securing the best possible data foundation is a crucial step in the investigative process. Potentially, this conceptualization of evidential standards could capture an initial phase in an extended process of evidential reasoning—evaluating whether there is an adequate data foundation to allow the investigation to move forward with specific evidential claims. The language used to report such data and the qualifiers (that is, modal operators; cf Toulmin, 1958) attached to the evidential claim should reflect the preliminary status of research (see Section 6 on communication). In this sense, this more charitable reading could be described as "compatibilist", as its endorsers would be primarily concerned with whether the collected data are compatible with the claim they seek to defend.

Let's turn to the second approach. The "scaffolding approach" interprets standards not only as guidelines for assessing support (that is, the data provided), but also as principles governing how warrants should be constructed, which background knowledge should be leveraged, what techniques should be employed and which alternative scenarios should be considered and how they should be ruled out. Simply put, on the scaffolding reading, evidential standards are standards about the construction of inferential scaffolds. These, as we said, should not be taken too rigidly; rather, they are situated, dynamic, and flexible, thus reflecting the norms of knowledge

construction of the scientific community in a particular historical and practical setting. Here, the normative force is placed on the robustness of the logical bridges between the data and the conclusion—how the inferential scaffolding is to be built for a target claim to be legitimately inferred.

This latter conceptualization carries an important implication. Even with all the "right" traces in place, evidential claims can be led astray in various ways. This can certainly happen through not engaging with adequate rebuttals and thus with relevant eliminative reasoning (Forber, 2011; Stanford, 2006). Additionally, even if the evidential claim is potentially correct (not false), it can be poorly inferred through inadequately constructed inferential scaffolding.

Crucially, a focus on scaffolds, warrants and rebuttals can feed back onto the support step of the inquiry, and thus influence data requirements. In particularly ambiguous contexts, the data collected may be insufficient to rule out potential defeaters, thereby promoting the hunt for additional or higher-quality data. Once again, under this account, standards are not carved in stone, but are both sensitive to the context of inquiry and transformed over time as new empirical findings and theoretical acquisitions come online, and debates evolve. As a consequence, this second conception of evidential standards leaves more reasonable room for disagreement: dissent in this case can "creep" into each of the numerous inferential steps that link data foundation to evidential claims, as well as arise from the evolution of standards over time. In this sense, standards so understood, while providing normative constraint on how inferential scaffolds should be constructed, do not represent a "magic bullet" that automatically resolves an interpretive problem at hand. At this point, it should be clear why the presence of explicit evidential standards cannot guarantee the settlement of interpretive problems. Even when agreement on what the standards are is in place, differences in their interpretation—whether in varying conceptualizations or application of the same conceptualization—could still be a source of epistemic tension.[20] As we will see, a scientific community can establish norms of alignment with the adopted conceptualization and implementation of explicit evidential standards, thus exposing failure of adherence. This, we believe, best describes the dynamics at play in the discussion on the 2023 preprints.

[20] Jacoby, referring to what he calls "epistemic standards" (such as a preference for more parsimonious explanations), effectively expresses how although a community may share the same standards there can still be disagreement: "[s]o though all scientists in a community endorse the same standards, they do not necessarily give the same credence to the same standards, nor do they necessarily apply the same standards in the same way (Jacoby 2023, p. 44)

As an example of this, consider how Berger et al (2023a) place greater emphasis on skeletal completeness and articulation, while Stiner (2017) focuses more on cultural aspects such as burial site and presence of artifacts.

Let us now situate the above analysis within the context of *H. naledi* burials. In what sense are the standards described by Martinón-Torres and colleagues taken to be met? The *H. naledi* team seems to best align with the data-checklist understanding of standards. This is already evident in the terminology used from the outset. The paper argues that "a combination of stratigraphic, anatomical, and taphonomic evidence supports this feature as a burial" and then lists the "five key observations", suggesting that individual standards should be viewed as data points to be collected, framing them in a model of empirical gathering. Far from being a merely linguistic choice, the paper can be seen as aimed at reporting descriptions and analyses *supporting* differences in sediment composition, stratigraphic disruption, anatomical coherence of skeletal remains and the matrix-supported position of some of these, and the compatibility of non-articulated material with decomposition and collapse. These, taken together, would support the evidential claim that it is a burial we're dealing with. There is certainly value in this endeavor, as surely any evidential claim is not raised in a vacuum, but it is grounded in the data collected. But support should not be conflated with warrant, as we've seen. The question then is whether this represents checklist thinking at its most problematic or most charitable.

While recurring use of expressions such as data being «compatible» (e.g. p. 13) or «consistent» (e.g. pp. 12-16) with burial claims may suggest the latter, the paper is ultimately not structured to promote a preliminary inquiry prioritizing data collection to support cautiously stated claims. It is, unfortunately, the most problematic understanding that is at play, as repeatedly raised in reviewer reports and commentators. Among key concerns, the claim that «the archaeothanatological reasoning is largely absent» (R2) and that "linking arguments" between data description and evidential claims are unclear (R3; see also Flint Dibble's public peer review) suggest the flattening of justification aspects onto those of supports. Here, the role of the evidential standards thus appears to be almost totally exhausted in the collection of relevant data, while the specifics of how the data collected warrant the conclusion and how possible rebuttals are eliminated seem to be comparatively neglected.[21] The anatomical analysis of the bones is a case in point, suggesting a form of reification of the third standard. While the team claims to identify some articulated body parts, their interpretation doesn't seem to adequately leverage background knowledge on comparative bone shifts in a matrix-supported context (burial) and in the open (R3), nor does it robustly contrast this interpretation with the possibility of accumulation by natural agents (Martinón-Torres et al., 2023). Thus, an inferential gap remains between the assessment of potential data support and its mediation as evidence for a burial claim.

[21] This is also evidenced in the very brief discussion section of the 2023 paper. There, the idea that the data collected can be interpreted as supporting the conclusion is more asserted than actively argued for.

What is worth noting, however, is that the arguments put forward by the *H. naledi* team appear to fall short also on checklist reading. Recall our previous discussion of quantitative and qualitative data inadequacy (Section 4.1): relevant data appear shadowed, lack resolution or are unfit for purpose. Yet by the lights of a data-checklist approach, securing a solid data foundation should be a central concern.

This line of criticism from reviewers and commentators part shouldn't be surprising as these—most obviously—adopt, and crucially enforce a scaffolding conceptualization of evidential standards. As mentioned, discussions of relevant warrants and rebuttals often feed back into the data itself, prompting a reassessment of what data is required to support the conclusion[22]. This is particularly relevant in chaotic and "messy" contexts such as caves, where the criteria must remain adaptable to the challenges imposed by the environment. In particular, the charges regarding the lack of primary analyses and the insufficient engagement with eliminative reasoning discussed in Section 5.1 imply readings of archaeothanatological standards as providing indications of how the inferential space should be navigated and constrained. Without objecting to the relevance of the identified standards, reviewer reports and related papers aim at enforcing their adequate implementation.

Let us take stock. Not only does the team fail to adhere to the evidentiary standards it claims to follow by misinterpreting them, thus adequately implement them within an inferential scaffolding framework. A failure of adherence is also evident on the more limited, data-checklist interpretation of evidential standards. Because these evidential standards carry normative force, it is through the peer review process that the paleoanthropological community enforces them and pushes for realignment in light of adherence failure. The public nature of this process allows us to observe these dynamics as they unfold and to assess whether responses genuinely move in the direction of realignment.

In this regard, in March 2025, a new updated preprint on burials was published in *eLife* (from September 2025 the official version of record). Although a detailed analysis of the differences between the two preprints is beyond the scope of the current chapter, a few aspects are worth noting (see also Chapter 2).

[22] To be fair, the *naledi* team incorporates at least one element attributable to the scaffolding approach, namely how general standards can be complemented by context-specific criteria. In Fuentes et al 2023 the team introduces three additional criteria, drawn from Pettitt (2022): the presence of artificial lighting, the existence of alternative entrances to the Dinaledi Chamber, and whether complete bodies (not just body parts) were deposited. According to the team the 2023 preprint fulfill the third criterion, while Elliot et al. (2021) reject the second, and the only support for the first comes from a non-academic source (Bower, 2022), with no peer-reviewed evidence of fire use by *H. naledi* to date.

5.3 Burials 2.0

In 2025, three new preprints were published, presenting updated data and analyses on *H. naledi*'s purported burials, engravings and their evolutionary implications. The preprints maintained the same overall structure, with the papers accompanied by the open reviews and the authors' response, but with a reduced number of reviewers: two each for the papers on burials and engravings (compared to four in the previous round) and only one for the paper on implications (compared to three previously). Nevertheless, as of September 2025, the preprints have been designated as *eLife*'s version of record, meaning they now constitute the definitive publications on *Homo naledi*'s meaning-making behaviors. Let's focus on the burial claims, in order to analyze the continuities and differences compared to the first preprint, and specifically on the following question: how, if at all, did the *H. naledi* team revise its interpretation of the evidential standards?

The new preprint reveals a shift in the declared argumentative strategy. Specifically, any explicit reference to the previous evidential standards has been removed from the new version, and Mtoto is no longer presented as an exemplary case on which to model those criteria. In their place, the *H. naledi* team advances a *minimal definition* of *primary cultural burial* (Berger et al. 2025). The definition consists of three criteria:

1. A hole or pit is dug by hominins into sediment.

2. A body or parts of bodies are placed into this feature by hominins.

3. The remains are covered (backfilled) by hominins (pp. 57-58).

This shift is justified by (i) an emphasis on the variability of interment practices in the Palaeolithic and more recent times (both in their original features and subsequent taphonomic transformations) and (ii) the aim of avoiding loaded requirements of culturally sophisticated expressions, such as grave goods or symbolic elements.[10] In this way, these criteria should apply across all funerary contexts, without excluding hominins with smaller cranial capacities whose behavioral complexity cannot be expected to match that of *Homo sapiens*[23].

According to the team, the hypothesis of cultural burial is testable by examining evidence for criteria (1)–(3). As in the first preprint, the article presents analyses, including new ones, that show how the deposits fulfil all three criteria. Some of the revisions implemented appear to address the

[23] The initial argumentative strategy can be reconstructed in terms of a parity of reasoning principle (we are grateful to Adrian Currie for pointing this out), according to which comparable cases should be assessed using the same inferential standards, unless independently justified asymmetries are introduced. In the first preprint, parity considerations were used to motivate the extension of *Homo sapiens* burial standards to *Homo naledi*, whereas in the final version the argument is inverted, with differences in the biological capacities of *H. naledi* and *H. sapiens* now taken to justify a divergence in the criteria applied.

concerns raised by previous reviewers and commentators (see the authors' response). For instance, the revised preprint/version of record now includes dedicated sections that discuss how both previously published data and new findings are used to test and rule out alternative hypotheses regarding the accumulation of bodies within the Rising Star Cave system. Additionally, the authors have incorporated some of the primary analyses missing from the first version, such as micromorphology. The conclusion remains unchanged: because the new analyses are consistent with the simplified criteria, the burial claim is taken to be adequately supported.

Two points are worth emphasizing. First, regarding point (ii), the presence of culturally salient elements did not appear fundamental to evaluating the burial hypothesis in the first phase of the debate, as the discussion focused primarily on the skeletal and sedimentological data (see Section 4.1). As such, this correction appears to misidentify the main reasons behind the voiced scepticism—the lack of adherence to invoked evidential standards—by attributing it primarily to a prejudice about the behavioral abilities of a small-brained hominin, which the revised definition is supposed to counter. Second, the lamented lack of adherence to evidential standards is addressed in the new version through the adoption of simpler, 'minimal' criteria, which would be better suited to accommodate the variability of interment practices mentioned in (i). However, if the initial disagreement did not concern the presented standards, their reformulation is epistemically interesting. What explains this change of strategy—from parity of reasoning to differential reasoning?[24] While we are in no way suggesting that revising one's view is inherently problematic in science—quite the opposite—we believe this move is again indicative of a persistent interpretation of evidential standards leaning towards the data-checklist approach.

Our reading is that the minimal definition achieves a maximisation of the fit of the criteria *to* the data. Particularly telling in this sense is that the new criteria intentionally leave open the possibility that sediment reworking could occur in cases of burial. The severe degree of disturbance of the presumed burials, with large displacement of bones and the intrusion of foreign skeletal elements, was indeed one of the most critical points highlighted by the reviewers of the first preprint. In the version of record, however, this phenomenon becomes perfectly compatible with the revised criteria. Indeed, a substantial portion of the new version is devoted to demonstrating that significant reworking of the deposits occurred, and that this is more consistent with hominins burying additional bodies or body parts than with natural processes. In this way, some of the most problematic aspects of the previous preprint become not only compatible with but even supportive

[24] This is not a straightforward question to answer. The new version was released recently, and reactions from the scientific community are still limited compared to the first preprint. As we've mentioned, there are also far fewer reviews, two instead of four. For these reasons, what we put forward here is a preliminary explanation that again calls into question the distinction between data checklist and scaffolding approaches to evidential standards.

of the team's hypothesis. Rather than addressing the misalignment with the community standards, the *H. naledi* team appears to have instead reiterated a data-checklist interpretation. In doing so, they adapted the standards to the data, consistent with an approach that emphasizes internal compatibility over the constrained construction of specific inferential scaffolds.

Nonetheless, the reception of the new preprint is evolving. One reviewer is now convinced that *H. naledi* intentionally buried their dead, while another remains uncertain, since concerns about how *H. naledi* got into the Rising Star cave system have yet to be adequately addressed. This state-of-the-art development, however, does challenge the claim by some of the involved authors that the burial hypotheses for *H. naledi* has finally gained acceptance within the scientific community[25].

These developments notwithstanding, there remains one point that has yet to be addressed, as noted in Section 4.1. In the following section, we will examine the communicative practices adopted by the team—which are most relevant to the early stage of the debate—and explore why they carry additional failures of adherence to community standards.

6. Communicating evidential reasoning in paleoanthropology

We have shown that the 2023 preprint on intentional burial in *H. naledi* failed to adhere to the standards that the funerary archaeological community endorses. Given the focus of our discussion so far, one might think that this failure is mainly epistemic. Until now, we have focused on reporting criticisms that concerned the "evidential reasoning" employed by the *H. naledi* team. However, it is the opinion of many reviewers and commentators that communicative strategies are hard to disentangle from evidential aspects. Communication, as we see it, occurs both within the community and from the community to the general public. In this section, we explore how the scaffolding conceptualization of evidential standards entails specific norms for science communication. If meeting evidential standards is primarily about adhering to norms in the construction of inferential scaffolds—the correct retrieval of data, the application of adequate methodologies and the use of eliminative reasoning in the face of acknowledged equifinality—then two key implications for science communication follow.

We term the first *inferential disclosure*: scientific communication should clearly convey the inferential scaffolds on which conclusions rest, along with their degree of stability, *without*

[25] See John Hawks' post on his blog: https://www.johnhawks.net/p/burials-by-homo-naledi. See also Jamie Hodgkins' (critical reviewer of both the first and second preprint) reply: https://www.youtube.com/watch?v=_Pc0J0TyFtk and also https://clas.ucdenver.edu/newsroom/2025/10/07/did-homo-naledi-bury-its-dead-or-not-where-do-we-stand

shadowing data and processes that might impede this assessment. This shifts the focus from merely reporting results to actively engaging with the underlying evidential reasoning and its associated profiles of uncertainty. The second is a *principle of proportionality*: the communicative strength, assertiveness, and rhetorical style used to present scientific claims should be proportionate to the stability of the inferential scaffolds supporting them. That is, the more uncertain, provisional, or underdetermined the inference, the more cautiously and transparently it should be communicated. Again, it is significant that communication surrounding the *Homo naledi* case—both within and outside the scholarly community—appears to depart from these principles.

First of all, there is a general lack of clarity that seems to have made it difficult for reviewers to evaluate the team's investigations. This lack of clarity impedes the assessment of relevant inferences, and thus affects "inferential disclosure". Many have exposed difficulties in interpreting the figures: for example, Figure 3b, which is supposed to show the stratigraphic profile of Feature 1, has a black square (intended to provide scale) covering much of the image. Thus, the basis on which stratigraphic disturbance is inferred is hard to assess. These issues affect not only the figures but also the body of the text in general. Violations of inferential disclosure emerge from the words of one of the reviewers on the matter: «[t]he paper is hard to follow. It lacks key citations, contextual background information to inform the reader about the geological and depositional structure of the caves, and concise understandable descriptions of the methods and the significance of the results (...) this evidence needs to be elaborated on the in text (*sic*), contextualized, and edited for clarity» (pp. 70-71).

Foecke et al. (2024) likewise focus on how the *H. naledi* team fails to meet the "reporting standards" in current scientific practice in geochemical studies. These standards cover not only the analysis and interpretation of data, but also their *presentation*. For example, the decision to present XRF and PSD results in non-standard formats makes it particularly difficult to assess the team's analysis according to the criteria normally applied in the relevant scientific community (see Foecke et al., 2024, section A). Likewise, in the presentation of PSD results, Foecke and colleagues complain about the lack of inclusion of the percentages of clay, silt and sand in the sediments, making it impossible to represent these results in the format that is normally standard in geosedimentology (a process implemented, among others, in Martinón-Torres et al., 2021). Once again, what is particularly problematic about the presentation of the data is the extent to which it undermines the transparency required for external evaluation of the inferences drawn.

Failure of adherence to the inferential disclosure principle also emerges from communicative choices obscuring the direction of reasoning. The extensive use of "persuasive communication devices" (Reviewer 4, Corneille et al., 2023) is reflected in the presentation of the research,

anticipating the conclusion and thus convincing the reader of its validity. For example, the features are introduced as burials at the beginning of the article, even before presenting any evidence in favor of this (2023a, p. 5). Another example of problematic communication is the case of the stone object "in the hand" of one of the individuals present in the Hill Antechamber Feature. This object, the first time it is mentioned in the text, is immediately presented as an "artifact", even before providing any type of analysis of it (p. 16). If not understood as a violation of basic communicative principles of transparency, the alternative is even more troubling: that the presentation accurately reflects the actual direction of reasoning adopted, with a narrative developed prior to the beginning of the fieldwork (see also Foecke et al., 2024).

Second, the principle of proportionality appears most glaringly violated by the fact that burial claims were announced through a massive media campaign before the scientific claims had undergone peer review. This is especially concerning given paleoanthropology's status as a "celebrity science" (Jones, 2019), which heightens the risk of sensationalism (Havstad, 2022) amid widespread public interest. Indeed, following the preprint's release, media coverage largely framed the naledi team's conclusions as definitive. A CNN article ran the headline "Mysterious species buried their dead and carved symbols 100,000 years before humans"; likewise, a post on the National Geographic website titled "New Evidence Suggests Intentional Burials and Use of Symbols by Other Hominins Hundreds of Thousands of Years Before *Homo sapiens*". The members of the *naledi* team, in turn, presented their findings to the media as conclusive, as exemplified by the Netflix documentary (see next paragraph).

This media framing is linked to a shared concern that communication choices and failure of adherence to evidential norms reflect haste in publishing and promoting a particular interpretation, despite the incompleteness of data and analyses (see the comments of the reviewer on the three papers published by the *H. naledi* team in 2023 and Pickering & Kgotleng, 2024). A key factor in this dynamic was played by *eLife*'s publication model, which allows the preprint to be made public, thus anticipating the standard publishing process (Pickering & Kgotleng, 2024). The near-concurrence of the release of the three articles, the related popular Netflix documentary, a popular book by Berger and the sending of *H. naledi* and *Australopithecus sediba's* remains into space further reinforces this interpretation. The end result is that the popular media broadcast the conclusions presented in preprints, the documentary and Berger's book, often without adequately contextualizing and accompanying them with the criticisms and doubts raised.

7. Adherence failures and responsibility

So far, we have argued that evidential standards in the paleoanthropological debate on *Homo naledi* are best understood as norms governing the construction of inferential scaffolds—from data to target claims. These standards are flexibly constructed and subject to iterative revision and refinement. Moreover, such evidential standards entail specific communicative norms, including inferential disclosure and a principle of proportionality between communicative style and the robustness of the inferential scaffolds. While a comprehensive analysis of how responsibility is assessed in paleoanthropological practice is beyond the scope of this chapter, we aim to draw a few important remarks from our case of interest.

Many philosophers have lately highlighted how scientists are responsible for considering the consequences of erring in their professional judgment. The concept of "inductive risk", lately revived in the work of Heather Douglas (2000, 2009), refers to the risks associated with judgments about the sufficiency of evidence required to make a particular scientific claim: they might err in making a claim in spite of sufficient evidence, or fail to make it in the presence of sufficient evidence. Scientists, differently put, ought to consider the predictable risks that accompany the possibility of mistake. As we've seen with the *H. naledi* case, archaeotanatological practice has come to gradually construct field-shared norms that are precisely meant to support evidential judgments. Differently put, we believe that evidential standards and the communicative standards that derive from them invoke a responsibility of adherence on the part of the members of the community. The ongoing public peer review process—and the degree to which correctives from peer review have been applied—makes this assessment equally in progress.

What is clear is that the media campaign and the failure to meet inferential transparency and proportionality principles represent a textbook example of what Havstad has termed "sensational science", with the potential of "amplified inductive risk"—where scientific results that gain significant public traction have been communicated in ways that irresponsibly amplify the inductive risks inherent in the study (Havstad, 2022). Havstad's case differs from ours in two important ways.

First, Havstad addresses the implications of pronouncements about the relative contributions of Denisovans to the genomes of present-day human populations, as well as the potential non-epistemic harms that may arise from errors in judgment—such as promoting notions of primitivism. Second, the inductive risk literature primarily focuses on the consequences of errors in professional judgment. Our concerns here, however, are somewhat different. While it might be argued that including a hominin species among those with burial practices does not carry harms comparable to those in Havstad's case, this does not mean that no harm is present. Many of these

risks are foreseeable and include reputational damage to the broader scientific community, as well as the misrepresentation of how historical sciences operate. Paleoanthropology, in particular, already contends with public perceptions of being little more than unconstrained speculation—even under the most rigorous and virtuous circumstances. Furthermore, while the question of whether *H. naledi* buried its dead may not have immediate societal implications, reputational harm and erosion of public trust in science can have broader, transversal effects—undermining the perception and trust for scientific research that *is* directly relevant to pressing contemporary concerns.

Crucially, our understanding of how evidential and communicative standards come into play carry one important implication: the above-mentioned risks persist *regardless* of whether the evidential claim will ultimately be accepted. Put differently, being right for the wrong reasons still carries significant consequences. It may well turn out that *Homo naledi* did bury their dead in an extremely hard-to-access cave system. However, the consequences of amplifying initial claims—especially those resting on widely regarded shaky inferential scaffolds—through a public media campaign will not be so easily undone. This should give us pause when evaluating strategies for communicating preliminary and disputed evidential reasoning.

8. Concluding remarks

This chapter has offered a philosophical analysis of the nature and role of evidential standards in paleoanthropology, drawing on an epistemically informative and widely discussed case study to highlight issues of broader theoretical interest. Our critical claims should not be taken out of context, nor should they be interpreted as denying the possibility that a small-brained hominin could engage in complex behavior, or as dismissing the significance of *Homo naledi'*s addition to the paleoanthropological record.

We have shown that, when it comes to evidential standards, fundamental disagreements may arise from differing conceptualizations of what these standards are intended to guide, how they should be met, and what implications they carry for science communication. We hope this contributes to a more attentive and reflective discussion—both within philosophy and human evolutionary studies—on the role and function of evidential standards, and how they help illuminate the epistemic profile and intended workings of the science of human origins.

SUPPLEMENTARY INFORMATION
IN-DEPTH OVERVIEW OF EVIDENCE FOR THE "KEY OBSERVATIONS" AND REVIEWERS' RESPONSES

Evidence in favor of the burial hypothesis

In this section, we will provide a detailed account of the evidence that the *naledi* team provides in support of the burial hypothesis, as well as the criticisms that the reviewers have raised. As mentioned in the main text, five "key observations" supports the idea that the three features are instances of primary burials (they actually attribute these remarks only to Dinaledi Feature 1, although then in fact they are implied in the treatment of the Hill Antechamber Feature):

1. the difference in sediment composition within the feature compared to surrounding sediment*;*
2. the disruption of stratigraphy;
3. the anatomical coherence of the skeletal remains;
4. the matrix-supported position of some skeletal elements;
5. the compatibility of non-articulated material with decomposition and subsequent collapse.

Evidence for 1 (The difference in sediment composition within the feature compared to surrounding sediment)

Petrography and geochemistry analysis [mainly x-ray diffraction (XRD) and x-ray fluorescence (XRF)] from within and surrounding the feature and comparison with sediments from other parts of the Rising Star Cave. Then PCA. «[T]he infill of the feature resembles the composition of other parts of the Dinaledi Subsystem and Lesedi Chamber where remains of *H. naledi* have been unearthed» (p. 7).

Analysis carried out only on Feature 1, not on Feature 2 and Hill Antechamber Feature. (Information is scattered between the main text, materials and methods and supplementary information, hard to follow).

Evidence for 2. Disruption of stratigraphy

Discontinuity of a laminated orange-red mud layer which gets fragmented and muddled at the level of Dinaledi Chamber Feature 1: «…the abundance of laminated orange-red mud clasts with grain sizes >2 cm around Feature 1 seems to be related to a laterally continuous layer where these clasts emanate (Fig. 3b). This layer … is continuous in the profile immediately to the east of the feature; it is disrupted in the sediment profile at the southern extent of the feature (Fig. 3b). Some orange-red clasts are visible within the exposed south face of the feature and isolated clasts occur within the feature itself» (pp. 6-7).

For Hill Antechamber Feature: the authors note that the floor of Hill Antechamber is composed of unlithified mud clast breccia, including laminated orange-red mud clasts, as in Dinaledi Chamber. They claim that the CT data show that these layers of clasts are interrupted within the feature, and that the latter exhibits two clast-rich zones. A bowl-shaped concave layer of clasts and sediment-free voids constitutes the bottom of the feature and at the south end of the feature this clast layer slopes in the opposite direction as the chamber floor. (p. 15). The CT scans are visible on p. 13.

Further evidence is described in *Supplementary Information 3* (pp. 41-42), but the section is not easy to follow.

Evidence for 3 (Anatomical coherence of skeletal remains)

For Dinaledi Feature 1: they state that «[d]uring this excavation, we *recognized* that the developing evidence was suggestive of a burial, due to the spatial configuration of the feature and the evidence that the excavated material *seemed* to come from a single body.» (p. 4, emphasis added) Skeletal remains have been identified and catalogued, but no further biochemical or taphonomic analyses have been conducted. Part of the feature remains unexcavated. (pp 30-40).

«The skeletal representation and spatial relationship of elements indicate that Feature 1 contains predominantly the remains of a single body including the eighty-three identifiable bone fragments and teeth that we recovered above and within the exposed circumference of the feature. Some of these were in direct contact with underlying fossil material, while others were above the feature separated by up to 5 cm of sediment. *Three of the fragments that we excavated directly above and in contact with Feature 1 represent two elements of at least one additional individual*» (p. 11, emphasis added).

For Hill Antechamber Feature: again, the main source of evidence is the CT scans, as the feature remains unexcavated. Through the scans they try to assess the spatial distribution of skeletal elements.

Evidence for 4 (the matrix-supported position of some skeletal elements)

For Feature 1: the fact that some skeletal elements (ribcage and cranial fragments) remain in a vertical or sub-vertical position (in particular near the edge of the feature) reflects the fact they were matrix-supported (and they cite Bolter et al., 2018; p. 12).

For Hill Antechamber: Berger and colleagues state that the vertical separation between the upper and the lower body is indicative of matrix support for the upper body. (p. 15).

Evidence for 5 (the compatibility of non-articulated material with decomposition and subsequent collapse)

For Dinaledi Feature 1: the *naledi* team does not seem to provide any real evidence in favor of this point. Rather, they seem to simply state it. «Without a water-saturated environment, the decomposition of soft tissue resulted in voids, into which the bones comprising anatomical units (head, torso, limbs) could collapse, resulting in the displacement of fragments and joint articulations. By contrast, the decomposition of a body within a natural depression on the cave floor results in a different pattern of displacement» (p. 13). This could be true, but there is no citation or argument that the pattern shown by the skeletal remains is compatible with the first hypothesis and not the second.

For Hill Antechamber Feature: the same situation as in Dinaledi Feature 1 is more or less repeated. There is some argument in favor of observation 5: «the maxillary dentition, with the mandible in or near articulation, is presently in a semi-inverted position, located above several rib shafts and near the right hand. This position is consistent with the displacement of the skull from a higher position, following decomposition of the nuchal musculature and cervical connective tissue, allowing the head to disarticulate from the axial skeleton, accompanied by rolling, inversion, and some horizontal displacement» (p. 15).

Reviewer responses to observations and related evidence

Key observation 1 (the difference in sediment composition within the feature compared to surrounding sediment):

- Reviewer 1: evidence needs to be elaborated in the text, contextualized, and edited for clarity.
- Reviewer 2: according to the reviewer, the geological information is detailed enough.
- Reviewer 3: It is not particularly informative that sediments associated with skeletal remains are different from sediments without bones. Analysis might suggest therefore not that these sediments were used to fill pits but that there has not been substantial reworking of the sediments. For a better understanding of the depositional context and the site formation processes the elemental analysis should be extended in a more systematic way after more extensive excavations.
- Reviewer 4: the authors took samples only from the Dinaledi Feature. The reviewer was not able to replicate the XRF PCA results. Also, there is an absence of data for both Na and S. There are visual comparisons of sediment grain size, shape, and composition, but these results do not distinguish samples from within the burial compared to those outside the burial, contrary to what the authors state.

Key observation 2 (the disruption of stratigraphy):

- Reviewer 1: evidence needs to be elaborated in the text, contextualized, and edited for clarity.
- Reviewer 2: according to the reviewer, the geological information is detailed enough.
- Reviewer 3: (Hill antechamber Feature) the team argues that the skeleton is lying more horizontally than the sediments, but there is no stratigraphic profile with the layers marked or fabric analysis. The three blocks should have been georeferenced, both *in situ* and in the lab. The items inside the blocks should be georeferenced as well.
- Reviewer 4: Figure 3b (Dinaledi Feature's south profile) is confused and there seems to be no difference in the profile that would suggest a disruption in the form of a pit. Also, no micromorphological analysis was conducted on this section to provide an evaluation of stratigraphic composition, and all the features remain partially unexcavated. Regarding the Hill Antechamber Feature, there are no visual or micromorphological data that demonstrate the bowl-shaped concave layer.

Key observation 3 (the anatomical coherence of the skeletal remains):

- Reviewer 1: articulated skeletal remains are found together with disarticulated remains of the same individual and others as well;
- Reviewer 2: there is no description of the relative sequence of joint disarticulation during decomposition and the spatial displacement of bones. A detailed assessment of the anatomical relationships of bones, both articulated and disarticulated, as well as the direction and extent of bone displacement, is missing. Figures are not clear enough.
- Reviewer 3: The authors provide no evidence of processes that could disarticulate the remains of unburied corpses, so the articulation of the remains recovered cannot alone be evidence of a burial.
- Reviewer 4: Figure 4 (specimens collected in relation to excavation area in Dinaledi Chamber) is not clear. No documentation (piece-plotting, 3D rendering of stages of excavation, etc.) of the elements that were removed from the Dinaledi Feature. Speaking of Hill Antechamber, skeletal remains show greater displacement than suggested by the authors. «There does not appear to be any intentional arrangement of limbs that may suggest symbolic orientation of the dead (another line of evidence often used to support intentional burial but omitted by the authors). Thus, skeletal cohesion is not enough evidence to support the hypothesis of an intentional burial» (p. 78).

Key observation 4 (the matrix-supported position of some skeletal elements):

- Reviewer 1: /
- Reviewer 2: /
- Reviewer 3: The literature on the decomposition of bodies is poorly cited and does not support the idea that some skeletal elements are matrix-supported. In addition, the patterns of dead body decomposition are much more complex than the authors would have you believe. The figures for Feature 1 are confused. For Hill Antechamber Feature, it is hard to understand the argument that the sediment separating the lower body from the upper body is evidence of burial. If the body was in an open pit it would have been flat with no separation. But there is no evidence of a pit here.
- Reviewer 4: The authors do not provide any evidence for 4.

Key observation 5 (the compatibility of non-articulated material with decomposition and subsequent collapse):

- Reviewer 1: /
- Reviewer 2: there is no description of the relative sequence of joint disarticulation during decomposition and the spatial displacement of bones. A detailed assessment of the anatomical relationships of bones, both articulated and disarticulated, as well as the direction and extent of bone displacement, is missing. The study does not address bone displacements within secondary voids created after the decomposition of soft tissues, nor does it provide assessments of the position of bones within or outside of the original body volume.
- Reviewer 3: The literature on the decomposition of bodies is poorly cited and does not support the idea that some skeletal elements are matrix-supported. In addition, the patterns of dead body decomposition are much more complex than the authors would have you believe.
- Reviewer 4: the reviewer was not able to understand the argument in favor of 5.

Summary of reviewers' positions

First reviewer: «In its current form the paper (…) does not meet the standards of our field» (p. 70). The first reviewer points out that the null hypothesis of the paper is that the features are instances of intentional burial and the authors seek support for the latter rather than testing it. The authors also considered only two hypotheses: either the features are cases of intentional burial or bodies left in natural pits. The most parsimonious hypothesis would be that the features are the outcome of geological phenomena or activities of carnivores. «To support this will take many more lines of evidence than presented here such as micromorphological analysis of the overall cave system and each feature (discussed in the supplementary information but briefly), full detailed reconstruction of sediment, water, fossil, and debris movement throughout the cave system coupled with reconstructions of body decomposition rates» (p. 72).

Second reviewer: «I feel that there is a significant amount of missing information in the study presented here, which fails to convince me that the human remains described represent primary burials … Insufficient evidence is provided to differentiate between natural processes and

intentional funerary practices» (p. 72). In particular, the reviewer thinks that the geological information is sufficient, but the archaeothanatological one does not.

Third reviewer: The main issue is that the purported burials are not fully excavated, so important lines of evidence are lacking. Also, there is a lack of detailed geological and archaeological evidence. «One of the key pieces of evidence for demonstrating deliberate burial is the recognition of a pit… That *H. naledi* buried their dead here can't be excluded based on the data, but neither is it supported here. My view is that this paper is premature and that more excavation and the use of geoarchaeological techniques (especially micromorphology) are required to sort this out» (p. 75).

Fourth reviewer: In addition to the meticulous point-by-point analysis of the 5 observations the reviewer states that the paper is written using language not suitable for a journal but more for a science news magazine. The authors also made extensive use of "persuasive communication devices" as illustrating the working hypothesis as an established fact before presenting any evidence. He is also the only reviewer who mentions the “artifact”. «Overall, there is not enough evidence to support the claim that *Homo naledi* intentionally buried their dead inside the Rising Star Cave system. Unfortunately, the manuscript in its current condition is deemed incomplete and inadequate and should not be viewed as finalized scholarship» (p. 78-79).

CHAPTER II
NO REST FOR THE DEAD: HOW TO TAME VARIATION IN HOMININ PALEOLITHIC BURIALS

1. Introduction[26]

In the previous chapter, we focused on a particularly heated debate in funerary archaeology: the claims surrounding the purported burials of *Homo naledi*. The debate also provided an opportunity to analyze the concept of evidential standards and how these standards arise, change, and are interpreted and used by the paleoanthropological community. In this chapter, we broaden the scope of our investigation to consider how scientists study Paleolithic burials more generally. We will pay particular attention to the variability observed in Paleolithic mortuary practices, focusing especially on the remarkable fact that at least two hominin species appear to have buried their dead. While we will return to the case of *Homo naledi* and the discussion of evidential standards, we will also examine other methodological strategies used to identify burials in the archaeological record. We will also analyze in greater detail the new argumentative strategy adopted by the *naledi* team in their recent preprint (from September 2025 the official version of record), which we briefly discussed in the previous chapter. But first, we will show how the problem of variability manifests in the archaeological record.

Let us begin with two iconic cases. Here we can see the so-called *Principe* (prince) of Arene Candide in Italy, a richly adorned Gravettian burial.

[26] Part of the content of this chapter was presented first at the workshop "*Perspectives on Human Variation: Philosophy meets Evolutionary Anthropology*" (Leuven, June 2025) and later at the workshop "*Problemi aperti in evoluzione umana. Prospettive dalla filosofia e dalla paleoantropologia*" (Urbino, October 2025). I am grateful to the participants of both workshops for their valuable feedback.

Figure 1. Principe of Arene Candide. Picture by Lorenzo Donzelli, Wikipedia, https://it.wikipedia.org/wiki/Caverna_delle_Arene_Candide#/media/File:Resti_del_principe_delle_arene_candide.jpg

The burial contains the skeleton of an adolescent male *Homo sapiens,* richly adorned, hence the appellation “Il Principe”. It was discovered at a depth of 6.70 meters within a bed of red ochre. The individual’s head was encircled by hundreds of perforated shells and deer canines, which likely formed a cap or headdress. The grave assemblage also included shells, mammoth ivory pendants, four perforated elk-antler *bâtons de commandement*—three of them decorated with fine radial striations around the perforation—and a 23 cm long flint blade placed in the right hand (Cardini, 1942; Sergi et al., 1974; Molari, 1994; Pettitt et al., 2003).

In stark contrast, the burial of Shanidar Z, a Neanderthal individual from Iraq, reveals a much simpler context. The body was recovered from what was perhaps a natural depression, with no clear evidence of grave goods or intentional preparation, apart from a triangular rock that originally was probably placed behind the head (Pomeroy et al., 2020a). The differing states of preservation between the two burials are also evident. Moreover, Shanidar Z belongs to a multiple-individual burial context.

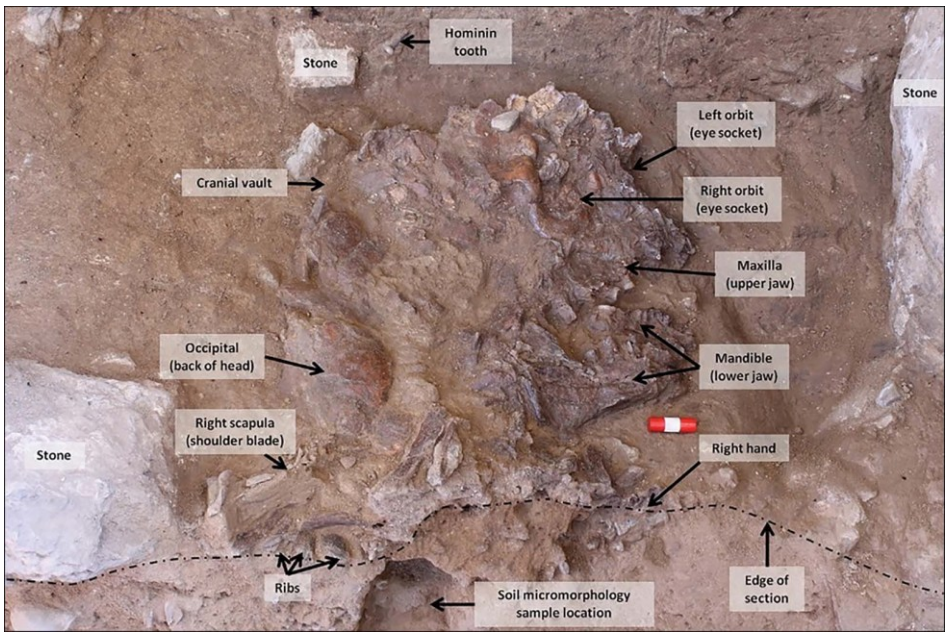

Figure 2. Shanidar Z. Pomeroy et al., 2020a

These two cases differ markedly: they belong to different species, distinct cultural traditions (Mousterian and Gravettian), and involve radically different treatments of the body. One is rich in grave goods and symbolic objects while the other is entirely devoid of them. To complicate matters further, they also exhibit different degrees of taphonomic preservation. While it is relatively straightforward to recognize *Il Principe* as a deliberate burial, the same cannot be said for Shanidar Z. Yet both have been classified as such. This raises a fundamental question: by what criteria do scientists identify burials in the archaeological record, and how do these criteria withstand the challenge posed by variability?

This chapter addresses the problem of cultural variability in burial practices as an epistemological challenge for the historical sciences. We begin by examining what is meant by "cultural variability" in relation to burials, showing how these vary systematically throughout the Paleolithic. In what follows, we examine how paleoanthropologists and funerary archaeologists define burials, arguing that the variability of these phenomena poses an identification problem analogous to that faced by evolutionary biologists. We then present the main strategies used to identify burials in the archaeological record. Finally, we offer a tentative assessment of these approaches, reflecting on their respective strengths and limitations when confronted with the pervasive variability of mortuary practices.

2. The variability in paleolithic burials: a review

During the Paleolithic hominins displayed a remarkable diversity in how they treated their dead. Intentional burials appear in the archaeological record only from the Middle Paleolithic onward, first in Western Asia (roughly after c. 120,000 years ago, Tillier, 2022), with earlier species engaging in other funerary practices such as *Cronos compulsions* (the dismembering and consuming of a dead body) and structured abandonment but not intentional burials (Pettitt, 2011). When Paleolithic hominins did bury their dead, the burials varied widely across several dimensions, from the number of individuals interred, to body positioning, grave goods, use of pigments, and the very locations and structures of graves. The following paragraphs aim to provide an overview of this variation in burial practices, with a particular focus on the Paleolithic (c. 3.3 million–11,700 years ago) and more specifically the Middle (c. 300–50 ka) and Upper Paleolithic (c. 50–12 ka). Both Neanderthal and *sapiens* sites will be presented, highlighting the systematic differences between the two, where they occur. At present, there does not appear to be convincing evidence that other hominin species buried their dead (see also Chapter 1). As far as possible, only sites whose intentional funerary nature is certain will be included. This review does not claim to be exhaustive, but simply to give an idea of the extensive variability that can be encountered in the study of Paleolithic burials. As far as we know, no systematic reviews of burial variability across the entire Paleolithic record have been attempted so far (but see Been & Barzilai, 2024, for a systematic comparison of Neanderthal and *sapiens* burials in Western Asia).

For the purposes of this overview, a broad working definition of burial will be adopted, one that includes all cases in which the body of a hominin was intentionally laid to rest by conspecifics. This provisional definition serves as a starting point, and we will return later to examine more closely the conceptual and methodological issues surrounding what qualifies as a burial.

2.1 Pits

Paleolithic hominin burials range from bodies placed in natural hollows to interments in deliberately excavated pits. The latter could vary in morphology, size, and depth. Artificial pits are usually recognized by sharp cut edges and fill sediment distinct from the surrounding matrix, while natural depressions often lack clear boundaries and may result from sedimentation or erosion (Pettitt, 2011). In some cases, natural depressions were intentionally modified to better accommodate the body; it even seems that bears' hibernation dens were sometimes used as burial niches for bodies, both in *sapiens* (e.g. the richly decorated Grotte de Cussac, Aujoulat et al., 2002; 2011) and in Neanderthals (possibly La Chapelle, Dibble et al., 2015).

Several well-known Neanderthal sites illustrate the difficulty of distinguishing intentional burials and pits from natural depositional events. For instance, at La Chapelle-aux-Saints a reanalysis confirmed a 40 cm deep anthropogenic pit with the Neanderthal skeleton intentionally covered soon after death (Rendu et al., 2014; but see Dibble et al., 2015). In contrast, the Roc de Marsal infant was once thought to lie in a dug grave, but later studies showed the deposition was natural (Sandgathe et al., 2011). Shanidar Cave illustrates both possibilities: some individuals may have fallen into hollows or been covered by natural processes, while others, such as the recently excavated Shanidar Z, show signs of purposeful burial (Pomeroy et al., 2020a). More definite examples of Neanderthal grave pits include La Ferrassie, where multiple adults and children were interred in prepared pits (Balzeau et al., 2020). In comparison, anatomically modern humans show a more consistent pattern of intentional burial in dug graves (see for example d'Errico & Blackwell, 2016; Martinón-Torres et al., 2021). In some cases it even appears that *Homo sapiens* corpses were buried in chert excavation pits, as in Nazlet Khater (Vermeersch, 2002).

2.2 Number of individuals: single vs multiple burials

Another fundamental dimension of variability in Paleolithic mortuary practices lies in the number of individuals interred. In both Neanderthals and early modern humans, single burials represent the predominant pattern. Yet there are notable exceptions, such as the double burial of an adult and a child at Qafzeh (90–100 ka, Tilllier, 2022), the Dolní Věstonice triple burial (c. 28 ka; Klima, 1987), and the Sungir double child burial (Alexeeva & Bader, 1998). Other sites, such as Shanidar Cave and La Ferrassie, reveal repeated use of the same location for successive burials, suggesting an emerging sense of designated mortuary areas. Mass interments are virtually unknown, though the Middle Pleistocene deposit at Sima de los Huesos (c. 430 ka, Bermúdez de Castro et al., 2004), containing at least 28 *H. heidelbergensis*/Neanderthal individuals, has been proposed as an early communal deposition. Taken together, these cases indicate that while single burials were the norm, multiple burials were possible as well.

2.3 Body positioning and arrangement of the body

Paleolithic burials display significant variability in the positioning of the body. A very common posture is the flexed or crouched position, often resembling a fetal curl, particularly characteristic of early modern humans in the Levant (e.g., Skhul and Qafzeh; Been & Barzilai, 2024). Neanderthal burials, by contrast, exhibit greater diversity, ranging from tightly contracted to fully extended supine positions. Examples include Kebara 2, laid supine with arms folded over

the chest and abdomen, and Shanidar 4, interred in a flexed posture. (Been & Barzilai, 2024). In several cases, Neanderthal burials also appear to be oriented along an east–west axis (Riel-Salvatore & Clark, 2001), though patterns are not uniform and may partly reflect taphonomic or other non-intentional factors. These variations in body position may have had cultural or symbolic significance, though their precise meaning remains difficult to reconstruct.

2.4 Grave structures and markers

Another key dimension of variability concerns the preparation of the grave itself. As we said, many Paleolithic burials were placed in shallow pits dug into the ground, a practice documented among both Neanderthals and early modern humans. Neanderthals sometimes employed stone supports—placing slabs beneath the head, bracketing the body with large rocks, or laying infants on beds of tools—as seen at Shanidar and Amud 1 or at La Ferrassie child burial, which was sealed by a slab (La Ferrassie 6, although whether the slab was intentionally placed is still controversial, Pettitt, 2011, p. 129). These choices might have held cultural significance, serving as markers, possibly symbolizing sleep or serving a practical purpose in wrapping the body. The use of props may likewise suggest, though not conclusively, an intentional concern with both the stability and the meaning of the placement. Early modern humans in the Levant, by contrast, generally lacked such features (Been & Barzilai, 2024), although they occasionally enhanced graves with other objects. For instance, at Krems-Wachtberg two *Homo sapiens* infants were covered with a mammoth scapula supported by a tusk. (Einwögerer et al., 2006). Burials were also associated with hearths, ochre spreads, or slight mounds. Overall, grave preparation varied greatly in scale and elaboration.

2.5 Grave goods

One of the most informative aspects of Paleolithic burials is the presence or absence of grave goods, which range from none to exceptionally rich assemblages. Most burials—particularly those of Neanderthals, but also many belonging to Upper Paleolithic *Homo sapiens*—contain few or no offerings. In some cases, the only objects found are utilitarian tools or animal bones, which may have been part of daily life or simply became incorporated into the burial infill along with the sediment used to cover the body. More elaborate practices, however, are attested. Both Neanderthals and early modern humans occasionally included animal remains such as jaws, antlers, or horn cores, as seen with Skhul 5, Qafzeh 11, and Teshik-Tash, although it is difficult to say if they have carried symbolic or ritual significance (Vandermeersch and Bar-Yosef, 2019; Been

& Barzilai, 2024). Upper Paleolithic humans were sometimes buried with ornaments made of shell beads and animal teeth, occasionally arranged as necklaces or sewn into clothing, as in the "Prince" from Arene Candide. In other cases, the grave goods included weapons, such as the mammoth ivory spears from the Sungir burials (Trinkaus & Buzhilova, 2018). These practices may suggest that, by this time, burials were clearly more than mere pragmatic disposals—they perhaps embodied elements of ritual, symbolism, and the expression of personal or social identity.

Neanderthal burials were once thought to include symbolic offerings, the most famous example being the alleged "flower burial" of Shanidar IV, where clumps of pollen were interpreted as evidence that flowers had been deliberately placed on the grave. Subsequent analyses, however, have cast serious doubt on this interpretation, showing that the pollen could have been introduced naturally by burrowing rodents rather than by human agency (Sommer, 1999). To date, there is no compelling evidence that Neanderthal burials included grave goods (Pettitt, 2011; Been & Barzilai, 2024). Other practices remain noteworthy: again, the deliberate placement of stones around or over a body, which appears in several Neanderthal burials and may reflect some symbolic meaning. Overall, while simple interments without grave goods were the norm, the archaeological record reveals a spectrum of practices, from minimal inclusion of everyday items to highly elaborate offerings.

2.6 Use of pigments

One of the most striking features of Upper Paleolithic burials is the use of pigments and personal adornments. Red ochre, often applied to the body, the grave, or grave goods, is documented in many *Homo sapiens* sites such as Border Cave (d'Errico and Backwell, 2016), Krems-Wachtberg, Lake Mungo (Bowler et al., 1970), Sunghir, and many more. The practice is generally interpreted as a symbolic treatment of the body. In contrast, although the use of ochre is attested in Neanderthal (Roebroeks et al., 2011), it is so far absent in burials, marking an important cultural difference.

2.7 Location of burials: caves, rock shelters, and open air

The location of Paleolithic burials varied considerably, shaped probably by both cultural choices and environmental conditions. Neanderthals often placed their dead deep inside caves, at least in the Levant, whereas early modern humans in the same region preferred cave entrances or rock shelters (Been & Barzilai, 2024). In the Upper Paleolithic open-air burials became more frequent, exemplified by sites such as Sungir (Trinkaus & Buzhilova, 2018). The first Neanderthal

remains from an open-air site have recently been found, although it remains unclear whether they were the result of a burial (Been et al., 2017). Rare cases, such as the decorated Grotte de Cussac in France, suggest caves used primarily as mortuary spaces, separate from habitation (Aujoulat et al., 2002; 2011). Some burials also occurred within settlement areas, as at Dolní Věstonice, hinting at a close integration of the dead within living spaces (Klima, 1987).

2.8 Demography: who was buried?

Not everyone in the Paleolithic received a formal burial, and the demographics of those interred suggest cultural selectivity. Adults, particularly those in their prime years, dominate the record, while infants and young children are underrepresented. Neanderthals appear to have buried infants more consistently than early modern humans (Stiner, 2017). Males are also more frequently represented than females, though notable elaborate female burials exist (e.g. Trinkaus & Jelínek, 1997). Certain individuals stand out for having received special treatment, whether through rich grave goods or careful burial, despite physical disabilities (e.g., Shanidar 1 and the two adolescents in the Sungir double burial), suggesting social recognition or care for the sick (Formicola, 2007). Importantly, formal burials remain rare overall, indicating that many individuals were treated differently and that preserved graves reflect only a selective portion of past populations.

2.9 Other mortuary practices: secondary treatments, cremation, and exposure

Beyond primary inhumation, Paleolithic mortuary behavior also included secondary treatment of the dead. At sites such as Grotte de Cussac, human remains were disarticulated, sorted, and redeposited, sometimes with selective emphasis on crania or long bones, hinting at complex rituals of secondary burial (Aujoulat et al., 2002; 2011). Similar practices are evident at Brillenhöhle (Magdalenian context), where bundles of bones were reburied, showing that remains could be intentionally handled long after death (Orschiedt, 2002). Other treatments included defleshing and cannibalism, documented at Neanderthal sites like Krapina, Croatia (Russell, 1997), and among Magdalenian populations, where skulls were shaped into cups (Marginedas et al., 2025). In addition, rare cases of cremation (e.g., Lake Mungo in Australia, Bowler et al., 1970) and probable excarnation suggest still further diversity. These practices demonstrate that “burial” was only one option within a wide mortuary repertoire: Paleolithic populations also processed, curated, or exposed the dead (see Pettitt for an overview of other mortuary practices during the Paleolithic).

To sum up, we have seen that there is great variability in Paleolithic burials:

- Pits: natural depressions, sometimes modified (behavior apparently more widespread among Neanderthals); artificial pits, also individual cases of bears' hibernation dens, chert excavation pits;
- Number of individuals: single burials were the norm, although multiple burials were possible as well. No mass interments known;
- Body positioning: extensive variability, *sapiens* seemed to prefer the flexed position;
- Grave structures and markers: employment of stone supports is present only in Neanderthal, although *sapiens* sometimes enhanced the grave by other means;
- Grave goods: absent in Neanderthal, wide variability in *sapiens*, from none to very rich grave goods;
- Use of pigments: absent in Neanderthal, while *sapiens* applied ochre on the body, grave goods and grave.
- Location of burials: at least in the Levant, Neanderthals preferred the deep interiors of caves, while *sapiens* privileged rock shelters and cave terraces. Open-air burials are attested as well, from the Upper Paleolithic;
- Demography: slight differences between Neanderthal and *sapiens* in the demographics of individuals buried. Burials of pathological individuals are attested in both species;
- Other mortuary practices: secondary treatments of the body is attested in Neanderthal, *sapiens* and even other hominins.

3. The variability in paleolithic burials: what problems does it pose?

From our review, we have seen that there is great variability in primary burials throughout the Paleolithic period, and this variability concerns several dimensions. Moreover, at least two species of *Homo* (Neanderthal and *sapiens*) are generally regarded as having practiced burial, with some systematic differences that seem to emerge from a preliminary review of the funerary evidence.

This variability is not merely descriptive; it has significant epistemic implications. As is by now well known, paleoanthropologists and archaeologists already face a variety of problems. They work in conditions of "epistemic scarcity" and "unlucky circumstances", with fragmentary data, disturbed contexts, preservation biases, and rampant problems of underdetermination (Currie 2018; 2021). In addition, they must grapple with the specific difficulty of reconstructing past systems of symbols and meanings (e.g. Kuhn & Stiner, 2007). But there is an additional challenge, what we may call the *variability-of-expression* challenge: the challenge of reliably identifying and attributing the same underlying cultural behavior in the archaeological record when its material

manifestations differ across space, time, and especially species, thereby putting our epistemic practices to the test.

This issue is particularly relevant in archaeology for two reasons. First, behaviors that vary across the record are intuitively more difficult to identify. Second, when the same behavior appears in different species, an additional question arises: how should the standards for recognizing that behavior be calibrated?

Regarding the first problem, intuitively, if the same behavior can manifest itself in different ways, this means that it may be more difficult to identify in the archaeological record, but this, of course, also depends on the behavior in question. Paleolithic burials, in particular, illustrate this problem, as it is often difficult to distinguish them from natural accumulation processes (see also Chapter 1). Their considerable variability further complicates identification, since relying on a fixed set of diagnostic criteria may not be sufficient.

But the problem raised by this variability is even deeper, and stems from the fact that several species seem to have buried their dead. As we will see shortly, claims of Neanderthal burials have given rise to debates in which the evidential bar for recognizing complex behaviors often appears to be set higher for non-sapiens species than for *Homo sapiens*, even if researchers rarely state this explicitly (see Meneganzin & Killin, 2024, for discussion of aesthetic capacities). The difficulty of establishing appropriate evidential standards for Paleolithic burials has also been increasingly acknowledged within the scientific community. Recent debates over the purported burials of *Homo naledi*, the discovery of Shanidar Z, and the reanalysis of several Neanderthal sites have all highlighted the problem of variability and the risks of calibrating standards primarily on *sapiens* burials (Berger et al., 2025; Pomeroy et al., 2020b; Dibble et al., 2015; Sandgathe et al., 2011). Indeed, several funerary archaeologists and paleoanthropologists had already noted this issue much earlier (e.g., Pettitt, 2002).

Now that the problem of variability in funerary archaeology has been characterized, we can turn to its implications for our epistemic practices. To do so, we will draw from recent work in the philosophy of paleoanthropology addressing character (or trait) individuation in physical anthropology (Meneganzin et al., 2024). This is a well-known problem in evolutionary biology, where the identification of authentic biological traits is a highly debated topic (see especially Wagner, 2001). Meneganzin and colleagues extend the discussion to human morphology, providing an overview of the different characterizations of the human chin. They propose a tripartite framework for understanding character individuation in paleoanthropology:

- Description: *what* biological structure is singled out as a meaningful character;

- Detection: *how* the character is measured in practice;
- Justification: *why* we are justified in treating the said biological structure as a single character.

We suggest that this same division can be applied to human behavior, as represented in the archaeological record, and is also useful for framing the problems raised by variability. Indeed, the variability-of-expression challenge raises issues both for the *description* of a behavioral trait (which actions represent instances of that behavior) and its *detection* (how to identify that behavior in the archaeological record). The *justification* for considering burial practices as a genuine trait will be left as an open question to be addressed in the future, although we will touch on this in the next section. We start with the description.

4. Back to basics: what is a burial?

In Section 2 we adopted a broad working definition of burial encompassing all the cases where a body was intentionally laid to rest by other hominins; now it is time to explore how variability poses problems for our description (i.e. definition) of burial. The Cambridge Dictionary defines a burial as «the act of putting a dead body into the ground, or the ceremony connected with this»[27]. Funerary archaeologists, however, use a more specific definition and tend to distinguish burials from other, closely related funerary practices. In an influential (and one of the few) books on Paleolithic burials, the funerary archaeologist Paul Pettitt provides a definition of burial which is also frequently adopted by other researchers: (see for example Pearson, 1999; Martinón-Torres et al., 2021; Berger et al., 2025):

> *Formal burial or inhumation*. The creation of an artificial place for the purposes of containing a corpse. This is at least a three-stage process involving 1) the excavation of an artificial pit or trench intended to serve as a grave; 2) the interment of a body within the grave; and 3) the covering of the body with the extracted sediment. Without the inclusion of humanly made grave goods, the result is part-natural (in that only natural phenomena are used) and part-artificial (as they have been repositioned). Formal burial may be distinguished from ritual deposits that include human remains, as with burial, the interment of the body itself is the prime object of the process (Pettitt, 2011, p. 9).

Inhumation/internment is different from funerary caching:

> *Funerary caching*. The structured deposition of a corpse, or parts of a corpse, in a chosen place, without modification of that place, such as at the back of caves, in natural fissures, etc. Also the use of pits originally created for purposes other than burial. Unlike structured abandonment, however, the place

[27] https://dictionary.cambridge.org/dictionary/english/burial

is given meaning beyond prosaic concerns such as corpse protection. Concepts such as "places of the dead" may arise from funerary caching (Ivi)[28]

The extensive variability outlined in Section 2 poses several challenges to such clear-cut definitions. As Pettitt himself acknowledges, several real-world cases do not fit neatly within any single definition. For instance, does placing a body in a naturally formed depression that has been modified by hominins to accommodate the corpse count as funerary caching or burial? And what if the body is placed in a bear's hibernation den? At first glance, this might appear to be a case of funerary caching, yet what if the act reflects a symbolic association between hibernating bears (which sometimes die during this harsh time) and the dead (Pettitt, 2011, referring to Grotte de Cussac)?

But variability also creates deeper disagreements. Some researchers do not distinguish between funerary caching and burial, using the latter term to encompass all cases in which a hominin was intentionally placed by conspecifics, whether in a natural depression or in a deliberately excavated pit (e.g., Rendu et al., 2014; Pomeroy et al., 2020a; 2020b). Other researchers, by contrast, place particular emphasis on the need to identify an artificially excavated pit (e.g., Gargett, 1989; 1999; Sandgathe, 2011; Dibble et al., 2015). As previously noted, Neanderthals appear to have been more accustomed to placing their dead in natural depressions. Perhaps not surprisingly, researchers who support the view that Neanderthals practiced burial tend to classify cases of funerary caching as burials, whereas those who are more skeptical of their cognitive capacities generally adopt a narrower definition. In these cases, the distinction between funerary caching and burial takes on particular importance, since for some researchers only the latter qualifies as a truly significant or complex behavior (e.g., Gargett 1989; 1999).

However, archaeologists rarely discuss the significance of distinguishing between funerary caching and burials. Pettitt, again, is an exception (2011). According to him, the two practices reflect distinct cognitive and communicative abilities and correspond to successive stages in the evolution of mortuary behavior. Burial would have evolved from funerary caching, first through the modification of natural depressions and later through the deliberate excavation of pits. In his view, this developmental trajectory would mark a major step towards symbolism (Ivi, p. 268; but see the Section on comparative and explanatory approaches for an alternative explanation). Consequently, proper burials would presuppose symbolic capacities and complex forms of

[28] Both funerary caching and formal burial differ from cairn covering, which represents an intermediate practice between the two (Pettitt, 2011, p. 9): «*Cairn covering*. The creation of a cairn – a pile of stones – over a corpse. This differs from funerary caching (whereby a corpse is taken to a natural feature and placed within it) as natural materials are brought to the corpse and used to cover it. The resulting space is part-natural, part-artificial, and in this sense similar to simple inhumation.»

communication, whereas funerary caching would not. Simply demonstrating that a body can be stored in a natural niche would suffice. By contrast, burials would entail «the communication of a set of ideas about a chaîne opératoire of three or more stages» (Pettitt, 2011, p. 268)—the excavation of a pit, the placement of a body, and the subsequent infilling—thus involving planning, coordination, and the communication «of an elaborate set of beliefs» (Ivi, p. 268). Pettitt remains one of the few scholars to provide a clear theoretical justification for this differentiation, and his framework suggests that future debates would benefit from a closer examination of the cognitive and social mechanisms underlying these diverse mortuary practices.

Thus, there are different opinions about the correct definition of burial. Variability first and foremost raises the problem of identifying the trait of interest (that is, *detection*, in Menegazin and colleagues' framework). Cultural and behavioral traits studied by archaeologists are no different: as in evolutionary biology, the first step should be to clearly identify the trait of interest[29].

These considerations have significant implications for scientific practice. As the following sections will show, the way a trait is defined inevitably shapes the way it is identified, an insight already emphasized by Meneganzin et al. (2024). For this reason, reaching consensus on the semantic extension of the behavioral/cultural trait we are studying is a fundamental starting point if we want to achieve consensus on the methods for identifying it in the archaeological record. It is to the *detection* of burials in the archaeological record that we now turn.

5. How to recognize burials in the archaeological record

Let us now turn to how the scientific literature has approached the problem of the identification of burials. What emerges are distinct *modes of inquiry*, each designed to focus on a different dimension of the investigation, or, in other words, to answer distinct research questions. Broadly speaking, paleoanthropological and archaeological inquiry unfolds through several stages: collecting relevant data, evaluating it as evidence for a specific claim, excluding alternative explanations, and so forth (see also Chapter 1, Section 5.2; and Toulmin, 1958). In funerary archaeology, at least four of these modes of inquiry can be identified: diagnostic lists of criteria, minimal definition, null hypothesis and complementary approaches. Since these modes of inquiry target different aspects of the scientific process, they are in principle compatible with one another, though, as we will discuss later, important tensions can nonetheless emerge in practice. We will

[29] In the literature, it is difficult to find explicit reasons for treating burials as a single behavioral trait. This may reflect a fundamental difference between evolutionary biology and archaeology: while the former requires the delineation of biologically meaningful traits, the latter can afford a certain degree of vagueness.

examine each of these strategies in turn and then assess them in light of the variability-of-expression challenge.

5.1 Diagnostic lists of criteria

We have already examined this strategy in detail in the first chapter, where it was the main focus of our analysis. Diagnostic lists of criteria (or "evidential standard", as the *naledi* team calls them) guide how traces are collected, configured as evidence, and used to justify specific claims. Criteria can be more or less flexible and adaptable to the context based on their interpretation (see Chapter 1), and include geochemical, taphonomic and anatomical evidence as well as body positioning, grave goods, and even the placement of the body in a place well-attended and important for the living (Stiner, 2017). This reliance on a plurality of criteria, rather than on any single indicator, constitutes a strategic move in the historical sciences: since individual strands of evidence are often defeasible, especially under the threat of equifinality, convergence across multiple sources provides stronger inferential support (Forber & Griffith, 2011; Perreault, 2019).

Almost every treatment of burials in the Paleolithic refers to some list of diagnostic criteria (e.g. Gargett, 1989; 1999; Pettitt 2011; Been et al., 2017; Pomeroy et al 2020b; Martinón-Torres et al., 2021; Berger et al 2023a and reviewes). Although there is no complete consensus regarding which criteria should be included in these lists, or on how they should be met (again, see Chapter 1), there is nonetheless broad agreement that some standards are more salient than others, such as the anatomical integrity of the body (especially the most fragile joints), the presence of grave goods and symbolic elements, and the excavation of a pit. To return to one of the cases mentioned in the second chapter, to support the claim that Mtoto, the Panga Ya Saidi child was indeed buried on the spot, the team provided the following justification:

> Four features demonstrate that a fresh body was placed in a location where the entire process of decomposition took place: (a) the macroscopic anatomical integrity of the body, especially unstable articulations; (b) the minimal displacement of bones, with movements explained as a consequence of decomposition; (c) the abundance of terrestrial gastropods that feed on earthworms in close proximity to the corpse; and (d) geochemical and histological analyses that indicate in situ decomposition and putrefaction processes. The PYS child remains meet all four of these *criteria* (pp. 96-97, emphasis added)

For Martinón-Torres and colleagues meeting these four criteria provides strong evidence that Mtoto was intentionally buried rather than simply abandoned. This example illustrates how diagnostic lists of criteria operate in practice: by establishing a structured evidential

framework that links observable traces to behavioral inferences under conditions of uncertainty.

5.2 The minimal definition approach[30]

A different perspective on diagnostic lists of criteria can also be adopted. Here we refer specifically to the *naledi* team's proposal in the second preprint (now the version of record) on burial evidence (see Chapter 1; Berger et al., 2025). Berger and colleagues introduce a minimal, testable definition of "cultural burial", understood as burial produced by hominin agency rather than natural processes. The definition consists of three criteria (echoing Petitt's definition):

1. A hole or pit is dug by hominins into sediment.

2. A body or parts of bodies are placed into this feature by hominins.

3. The remains are covered (backfilled) by hominins (p. 66).

The *naledi* team claims to adopt this parsimonious, process-based definition to avoid relying on culture-specific diagnostics (grave goods, symbolic objects, ritualized positions, "special" settings, etc.) that are often absent even in well-accepted human burials and are unlikely to be universal across hominins. The *naledi* team formulated this definition with the problem of variability in mind: given the broad diversity of mortuary practices, researchers should prioritize skeletal and sedimentary data. Although not included in the definition itself, the presence of grave goods and other criteria, such as differing degrees of preservation between animal and hominin bones, may nonetheless act as additional standards and strengthen the hypothesis of intentional burial (pp. 74-76).

According to the team, the hypothesis of cultural burial is testable by examining evidence for criteria (1)–(3). *Prima facie*, two interpretations of these criteria are possible: as necessary and sufficient conditions and as sufficient conditions only. The *naledi* team appears to interpret the criteria as necessary and sufficient conditions, at least in the context of the Rising Star cave system: all three must be satisfied in order to establish that the bodies under study were buried, and evidence contradicting even one of them is taken to support the opposite conclusion («[i]f evidence proves that one of these three actions did not occur, then the hypothesis is falsified», p. 66). Yet a

[30] This is the expression used by the *naledi* team, but we wish to point out that it is not as minimal as it appears, since it effectively excludes several instances of Neanderthal behavior where no artificial pit is present. Also there are some examples of *sapiens* individuals buried in natural depressions such as bear hibernation nests (Grotte de Cussac, France). The decision to use the term "minimal" is likely linked to the choice to leave out cultural and symbolic criteria, perceived as more complex compared to anatomical and stratigraphic standards.

weaker interpretation is also possible, one that treats the criteria as only sufficient and is more in line with the scaffolding approach presented in Chapter 1. In this view, the three standards on the list are enough to demonstrate the presence of an intentional burial, without taking the absence of evidence for one of them as evidence of the absence of a burial. The two interpretations may also differ in their assessment of the usefulness of additional criteria such as the presence of grave goods, the degree of skeletal articulation, or the special nature of the place where the body was found. In the case of the first interpretation, additional evidence can only reinforce the conclusion once the minimal definition has been satisfied, but cannot by itself prove the presence of a burial. Under the second, by contrast, complementary lines of evidence may compensate for partial satisfaction of the minimal criteria—e.g., if (1) and (2) are well supported but (3) is obscured by taphonomic processes—and can justifiably raise confidence in the burial hypothesis.

Furthermore, as in the case of the list of diagnostic criteria, it is possible to be more or less flexible in the application of the minimal definition. This minimal definition could be considered valid for all funerary contexts, or it could be considered limited to certain species, periods, or sites. This second interpretation appears to be the one adopted by the *naledi* team, which stresses that, especially in the case of Neanderthal burials, a pit is often absent or difficult to identify (p. 75). For this reason, the minimal definition should also be adaptable to these cases, eliminating the requirement that hominins dug a pit[31].

5.3 Null hypotheses

While in experimental statistics the null hypothesis (H_0) refers to the assumption that no relationship exists between independent and dependent variables, in funeral archaeology the term has been more broadly adapted to indicate a default assumption of non-intentionality (see Chapter 3 on null hypotheses in cognitive archaeology). In this context, this strategy assumes that any deposit under study is the result of natural accumulation unless demonstrated otherwise; and therefore all hypotheses of natural accumulation must be ruled out before intentional deposition can be proposed. The justification for the adoption of this strategy is that the hypothesis of natural accumulation is also the most parsimonious, presumably because hominin agency is seen as a more complex explanatory factor. It is thus regarded as the preferred starting point. In practice, this often

[31] Another difference compared with diagnostic lists is that while the latter usually provides empirical indicators—stratigraphic, anatomical, and taphonomic traces—that can support the inference of a burial, the minimal definition instead specifies what must have occurred for a cultural burial to take place, emphasizing the role of hominin intentionality. In this respect, the definition deliberately leaves open the question of which material indicators constitute evidence of a burial in the archaeological record: it mentions a pit or hole dug by a hominin but makes no reference to sediment composition or stratigraphic discontinuities (as instead Berger et al., 2023a did).

entails applying a higher evidential bar to taxa other than *Homo sapiens* (Belfer-Cohen & Hovers, 1992). Indeed, it has been applied to Neanderthals and *Homo naledi* (Gargett et al., 1989, 1999; Martinón-Torres et al., 2023) while it is almost always absent in cases involving *Homo sapiens*.

Notoriously, this approach has been advocated by Robert H. Gargett in a series of skeptical works on the evidence of Neanderthal burials: «[i]t is not enough to say that humans *could have* produced a given deposit; it must be shown that nature could not» (Gargett, 1989. p. 161, emphasis in original), and «[t]he primary critical stance I assume in this work is that natural processes must be ruled out before invoking human or hominid behavior to explain archaeological occurrences» (Gargett, 1999, p. 31). More recently the reassessment of some Neanderthal supposed burial sites has motivated the use of null hypotheses. For example, in their analysis of the Roc de Marsal child, Sandgathe and colleagues (2011) adopt a list of diagnostic criteria, but at the same time they claim to place themselves among those who believe that «in the absence of indisputable anthropogenic indications, the potential for instances of naturally buried human remains to mimic intentional interments is great enough that *natural explanations must first be completely ruled out* before an anthropogenic one can be accepted» (p. 243, emphasis added).

This was also the approach taken by some reviewers and commentators on the first preprint on *Homo naledi* burials. For example, Reviewer 1 says that «[t]he starting null hypothesis should be that the bodies were naturally covered in sediment. Intentional burial requires extraordinary circumstances and requires multiple lines of solid evidence to support the hypothesis» (p. 71). Similarly, Martinón-Torres et al. (2024) in their critique of the claims of the *naledi* team state that

> A fundamental problem, also unanimously outlined by the peer reviewers of the three papers, is that the authors have chosen a *null hypothesis* that is the less parsimonious one. The null and alternate hypothesis procedure aims to challenge current conventional thinking, i.e., complex funerary behaviors are associated with large-brained species. Surely the finding of articulated bodies is compatible with both scenarios (natural and deliberate accumulation) but the hypothesis that needs to be refuted is the more parsimonious one (natural) (p. 7, emphasis added)

Interestingly, Martinón-Torres and colleagues did not follow the same approach in their article on Mtoto, the Panga Ya Saidi child (2021).

5.4 Other methods of inquiry: comparative and explanatory approaches

Although not involved in the collection and evaluation of evidence, there are modes of inquiry that can nevertheless contribute to the study of Paleolithic burials. The first may be described as a *comparative* approach. It focuses on identifying differences between hominin populations, whether coexisting within the same region or separated across time and space, or within the same population over time. As this is still a relatively marginal line of research, it has not yet acquired an established label; however, we will soon examine an example of a study employing precisely this method. The aim of comparing different populations is not to determine which is the most "advanced," but to explore their differences and assess whether any meaningful patterns emerge. In this sense, the comparative approach explicitly recognizes and embraces the diversity of burial practices.

This approach was recently adopted by Been and Barzilai (2024). The paper compares the Middle Paleolithic burial practices of Neanderthals and early *Homo sapiens* in Western Asia. The authors performed a systematic review of the two populations' burials, to test whether funerary behavior differentiates the populations. They found numerous common features but also several differences (see also Hovers et al., 1995):

- Use of different caves and of different parts of the caves: Neanderthals tended to inter individuals inside caves, while *Homo sapiens* favored cave terraces/rock shelters;
- Presence/absence of positional markers: Neanderthals often used large stones or "modified limestone pieces" as special positional markers near/under the head, while *sapiens* did not;
- Different body positions: *sapiens*' bodies were usually in a tightly flexed posture, while Neanderthals showed much more variation;
- Symbolic behavior: only in *sapiens* (ochre and marine shells).

From their survey and from datings of burials in Africa and Europe, they also infer a tentative reconstruction of the chronology and causes of this behavior: burials were probably innovated in the Levant, as a result of high demographic pressure between Neanderthal and *sapiens*, and from there spread in Europe (see Belfer-Cohen & Bar-Yosef, 2000, for a similar diagnosis of the increase of burial practices during the Natufian).

A further and widely employed approach in archaeology is the use of *ethnographic analogy*, whereby observations from contemporary or historically documented societies are used to interpret

past behaviors. When applied to mortuary contexts, this method involves drawing ethnographic comparisons across different populations to account for specific aspects of their funerary practices[32]. For this reason, it can be described as *explanatory*. For example, Pettitt (2000) suggested that the small number of Neanderthal burials is not only due to a taphonomic bias but likely reflects their harsh living conditions, mobility, and orientation toward the present, with individuals often left where they died rather than transported to designated places. A useful comparison according to Hovers and Belfer-Cohen (2013) is the Pirahã of the Amazon, whose culture and language are centered on immediacy and lack abstract notions of time or memory beyond a few generations. Accordingly, their burials are simple and pragmatic: the deceased is placed in a shallow grave, sometimes with a few belongings, covered with leaves and earth, without ritual or mourning. The graves, often dug on riverbanks, are soon eroded and disappear. Although the Pirahã are not direct analogs for Neanderthals, their practices show that minimal and perishable forms of burial are compatible with a cultural system focused on the here and now, offering a modern case of variability within which Middle Paleolithic practices can also be situated. This could provide an alternative explanation for the greater prevalence of funerary caching among Neanderthals. This would not be due to a lack of symbolic capacity, or not only that, but to a particular lived experience (*contra* Pettitt, 2011). Other examples are present in the literature (see Hall & Binneman, 1987, and their use of the institution of hxaro among the San to explain variability in grave goods from Later Stone Age burials in South Africa)[33].

Insights derived from these complementary approaches can substantially enhance our ability to recognize and interpret burials in the archaeological record. Comparative analyses highlight the diversity of practices across time, space, and *taxa*, helping us to discern whether any meaningful pattern emerges. In turn, ethnographic analogy provides a framework for understanding why certain forms of variability occur, as in the case of the Pirahã and Neanderthals. Together, these methods furnish expectations and interpretive models that can guide archaeological reasoning. In practice, this means that when we encounter new or ambiguous finds—say, a possible Neanderthal interment in the Levant—we are better equipped to evaluate the evidence against a range of plausible scenarios (more on this later).

[32] We are not using the term "population" here because we are implying that Neanderthals and *sapiens* are two metapopulations belonging to the same species, but only because these ethnographic analyses may also concern two different cultures of the same species (e.g. Hall & Binneman, 1987). Incidentally, I believe that Neanderthals and *sapiens* definitely belonged to two different species, although they were able to mate and produce fertile offspring (see Meneganzin & Bernardi, 2023; Meneganzin & Stringer, 2024).

[33] As is well known, the use of ethnographic analogy in archaeology is particularly insidious. We do not have space to explore this issue in depth, but Wylie (1985) and Currie (2016) provide two valuable philosophical introductions to the problem.

6. Interlude

Before evaluating the different methodologies involved in the recognition of paleolithic burials, a pause is necessary. We have presented the different modes of inquiry as if they were distinct, without providing sufficient justification for treating them as such. In fact, we already mentioned in the previous section that these different strategies focus on distinct aspects of the investigation process. In this section, we show how this is implemented in funerary archaeology.

The only methods of inquiry that clearly address the same aspect of the investigation process (albeit differently) are the list of diagnostic criteria and the minimal definition. In fact, both provide guidance on what data counts as evidence of primary burial, and, depending on the interpretation of the criteria, ensure that the conclusion is adequately supported. However, we have seen that the *naledi* team explicitly introduces this minimal definition in contrast to lists of criteria that include culturally complex elements such as grave goods and ritual body positions. Nonetheless, in many cases, the lists of criteria do not include these complex cultural standards, focusing more on stratigraphic, geoarchaeological, and anatomical data (see, again, Martinón-Torres et al., 2021; pp. 96-97). Cultural criteria, even when present, are generally not deemed essential, as the absence of certain grave goods or other ritual elements does not rule out intentional burial. Therefore, the distance between the two approaches is actually quite small. Moreover, if the minimal definition is interpreted as sufficient but not necessary and flexible, it can be regarded as a particular instance of a diagnostic list, effectively eliminating the distinction between them.

The adoption of null hypotheses, on the other hand, impacts eliminativist reasoning and the kinds of alternative hypotheses that must be considered, effectively raising the evidential bar that burial hypotheses must meet. As discussed in Chapter 1, a focus on rebuttals can, in turn, feedback on the data required to substantiate a particular claim[34]. In particular, this happens with species other than *sapiens*. The result is that part of the paleoanthropological community assumes that claims for burial (and more generally for complex behavior) require especially strong evidence in the case of other hominin species. Gargett (1999) is particularly instructive in this regard. The article proposes five questions (effectively functioning as diagnostic criteria) for identifying Neanderthal burials. The first asks: «What constitutes evidence of purposeful protection of the corpse?». Gargett's answer is that «unless a new stratum can be distinguished, there is no logical way to argue that the remains were purposely protected» (p. 33; see also Dibble et al., 2015). This

[34] One might argue that lists of criteria and null hypotheses are not truly distinct methods of inquiry, since both can bear on eliminativist reasoning. That is correct, although we do not believe there should be a clear-cut division between the various methodological approaches. The difference, as we conceive it, is rather one of emphasis: diagnostic lists focus on the kinds of type-traces to be collected, whereas null hypotheses emphasize the height of the evidential bar. Moreover, these methods are often presented separately by funerary archaeologists, and may even be framed in opposition (see the main body of the text).

sets an unrealistic standard, since in most cases a burial pit will be refilled with the same sediment that was excavated to create it. Requiring that a distinct layer be detectable is thus an overly stringent condition, one that even some *Homo sapiens* burials would fail to meet (Pettitt, 2011).

The other methods of inquiry on the other hand, the comparative and the ethnographic, serve different epistemic purposes, *classification* and *explanation.* As such, these approaches are clearly compatible with the others. Once again, the goal is to obtain a more accurate representation of how hominin burial practices vary (classification), and to try to provide an answer as to why those practices took one form rather than another (explanation), rather than determining how to recognize them in the archeological record.

Finally, methods of inquiry can be combined with each other, as they deal with different aspects of the investigation process. The null hypothesis approach is usually paired with a list of diagnostic criteria (see for example Gargett, 1989, 1999; reviewers of Berger et al., 2023a). In these cases, the list of criteria provides guidance on which data should be collected, while the null hypothesis sets how high the evidentiary bar should be, especially in cases of *taxa* other than *sapiens*. More recently, some researchers working on *naledi* and Neanderthal endorsed the use of a list of diagnostic criteria without adopting any null hypothesis of natural accumulation, as the latter approach tends to penalize other hominin species (see Pomeroy et al., 2020b; Berger et al., 2025). In these cases, the preferred approach is an evidence-based one and involves comparing the different hypotheses and choosing the one that better explains the available evidence (*inference to the best explanation,* see also Chapter 3).

7. A Critical Evaluation

We have shown that burial practices among Paleolithic hominins display considerable variability, with at least two distinct species engaging in the intentional deposition of their dead. This variability gives rise to what we have called the variability-of-expression challenge: the difficulty of reliably identifying and attributing the same underlying cultural behavior (burial) in the archaeological record when its material expressions differ across contexts, species, and time periods. Such diversity complicates the task of defining what qualifies as a burial and what does not. We have also examined several methods of inquiry employed by paleoanthropologists and archaeothanatologists to investigate these early burials, highlighting how each serves a distinct epistemic purpose. In the following section, we will address a central question: should the variability of a cultural phenomenon influence the tools we use to study it? Our answer will be

affirmative. To support this claim, we will offer a tentative evaluation of these different approaches in light of the variability-of-expression challenge.

7.1 Diagnostic list of criteria

How well this approach fares depends for the most part on how it is interpreted. If we adopt a data-checklist model, or, worse, one grounded in intuitive folk epistemology (see Chapter 1), we risk relying on tools that lack the necessary precision for the task. Such approaches tend to be rigid, with evidential standards that are not easily adaptable to different contexts. Burial practices vary not only across sites but also between *Homo sapiens* and Neanderthals; criteria that work well in one case may fail in another. By contrast, adopting what we have called a *scaffolding* interpretation allows for greater flexibility in addressing the variability-of-expression challenge. This approach treats diagnostic criteria as context-sensitive tools that can be adjusted to specific situations. For example, by not requiring putative Neanderthal burials to display clearly artificial pits. Determining how these standards should be adapted, however, remains a major challenge, and one that is linked with the variability in burial practices. It is therefore crucial to establish the principles that guide such adjustments in relation to each site under investigation. Although this complex issue warrants a dedicated discussion, the final section of this chapter will show how comparative and explanatory methods can contribute to this task, at least in certain contexts.

7.2 Minimal definition

As with the list of diagnostic criteria, the evaluation of the minimal definition depends largely on how it is interpreted. When read in a "weak" sense, the minimal definition effectively collapses into the diagnostic list, a case we have already discussed. We will therefore turn to its "strong" interpretation, which treats the criteria in the definition as both necessary and sufficient. Although excluding the presence of grave goods or symbolic artifacts may be a step toward adopting more equitable standards for other hominin species, two objections can be raised against this approach, one *ontological* and the other *epistemological*.

On the ontological level, not all burials involve the presence of a pit. It is worth recalling that Paleolithic burials display considerable variability in body placement, ranging from deliberately dug pits to natural depressions, some of which show evidence of modification. Requiring the presence of an artificial pit excludes all the other cases. But it is at the level of epistemology that the problem becomes more serious. Even if the burial in question originally involved a pit dug by hominins, this does not necessarily mean that the traces of such activity remain discernible today.

Time is usually a destructive agent in historical sciences, and it is not unlikely that it will erase all stratigraphic traces of a pit. In such contexts, relying on other types of evidence (such as cultural evidence) may be the most reasonable thing to do. Indeed, there are actual cases where a burial is not recognized due to the identification of an artificial pit (Riel-Salvatore & Clark, 2001; Federica Fontana, personal communication). In a harsh epistemic situation such as the one faced by funerary archaeology variability urges not to rely on a fixed definition, but to adapt the tools to the context.

7.3 Null hypothesis and *sapiens*-derived criteria

Null hypotheses do not fare well in the face of variation. Eliminating *all* hypotheses of natural accumulation is too strong a constraint, especially in the epistemic context of the historical sciences. A more reasonable strategy would involve delimiting relevant alternatives. In practice, however, this approach often serves less to promote epistemic caution than to deny other species the capacity for complex cultural behaviors[35]. Indeed, this approach has been opposed in recent years by those involved in funerary archaeology (e.g. Berger et al., 2025a; Pomeroy et al., 2020b).

The variability-of-expression issue is particularly relevant here. When applied to species like Neanderthals, this framework assumes that intentional burial should replicate the features typically associated with *Homo sapiens* burials, such as deep, clearly delimited pits. As a result, the absence of these elements is interpreted not as a different expression of the same underlying practice, but as evidence against intentionality altogether. In other words, the null hypothesis implicitly enforces a *sapiens-centric* benchmark of intentional burial, ignoring the possibility that different species could have expressed the same behavioral category through materially distinct means. The situation becomes even more problematic when null-hypothesis reasoning leads to the adoption of stricter criteria than those normally applied to *Homo sapiens*. In such cases, the approach runs counter to the very variability observed in the archaeological record, making it virtually impossible to demonstrate that other species buried their dead.

[35] At the Leuven workshop on *Human Variation*, one participant suggested that null hypotheses can be particularly useful when dealing with 'first cases'. In situations where a possible burial predates the earliest confirmed instance of intentional burial, a null hypothesis of natural accumulation should be adopted as a precautionary measure. After all, extraordinary claims demand extraordinary evidence. Although this would be a reasonable approach if properly regulated, in practice it is not how it works. In fact, if we look at *sapiens* early African burials such as the Panga Ya Saidi child, we see that this approach was not adopted, although it is a case that clearly backdated the earliest Paleolithic burial (at least in Africa).

7.4 Comparative and explanatory approaches

These last two methods of inquiry are very promising in terms of investigating the variability of Paleolithic burials, with some caveats. In particular, the comparative approach works well if there is a correlation between different variables. If there were no systematic correlation between particular populations and specific differences in burial practices, then a comparative approach would reveal no meaningful patterns. Put differently, if variation were entirely stochastic, there would be nothing to explain, nor any basis for predicting the form that a given burial might take. Luckily, this is not the case with paleolithic burials, at least in some cases: we know for example from Been and Barzilai (2024) that Neanderthals' burials from Western Asia tend to have some characteristics that differentiate them from *sapiens* burials, such as the placement of boulders to mark the burial site. Therefore, it may be useful to pursue a systematic analysis of Paleolithic and even later burials, with the aim of identifying both recurring and unique elements in the mortuary practices of different populations over time. The ethnographic, or explanatory, method is more difficult to apply, since using ethnographic analogies to infer the causes of deep cultural practices is notoriously challenging. Nonetheless, it remains the only approach that actively exploits variability to explain past funerary behaviors.

Moreover, these methods could work particularly well in combination with the lists of diagnostic criteria. The comparative approach could refine such lists by testing which criteria genuinely co-occur in different populations or periods, highlighting sets of traits that may function as robust indicators. For example, the recurrent placement of boulders over Neanderthal burials in Western Asia (Been & Barzilai, 2024) gains diagnostic strength once the comparative method confirms its patterned association with other features.

In turn, the explanatory approach contributes by suggesting what could be the reasons behind certain cultural practices. As we discussed, the apparent Neanderthal preference for burial in natural niches can be interpreted through ethnographic analogy. Rather than considering it as evidence of cognitive inferiority compared to *sapiens* that intentionally dug pits, as some reconstructions have implied, this practice may instead reflect the distinct ecological and behavioral niches occupied by the two species. In this way, the interaction between comparative and explanatory inquiry can actively improve the lists of diagnostic criteria, transforming them from static checklists into evolving instruments that better capture the complexity of Paleolithic mortuary variability. Considering the different modes of inquiry together provides a clearer picture of their potential synergies or incompatibilities[36].

[36] It is important to bear in mind that even these methods may suffer from excessive rigidity. For example, if we infer from Been and Barzilai (2024) that *all* Neanderthal burials in Western Asia must feature boulders delimiting the

8. Conclusions

To conclude: variability in burial behavior should not be viewed merely as noise, but as a feature of the phenomenon itself. Approaches that are flexible and context-sensitive, such as diagnostic lists of criteria, appear to be more effective. And when supplemented by complementary methods of inquiry, such as comparative and explanatory approaches, they offer an epistemically richer and more powerful framework. This allows us not only to detect burial behavior more accurately, but to better understand its diversity across time, geography, and hominin species. This last dimension—interspecific variability—is particularly significant from an epistemological standpoint. As we have seen, the possibility that other species buried their dead still causes heated debate within the scientific community. The challenge lies in how to investigate complex behaviors in other hominin species without either assessing them through criteria developed specifically for *Homo sapiens* or, conversely, adopting standards that are too permissive or idiosyncratic. As shown in Section 4, establishing a consensus on what constitutes burial and what does not, and why, would already represent a step forward, as it would provide a starting point from which to rethink the criteria for identifying burials.

In conclusion, variability in burial practices challenges our classificatory tools. A rigid typology cannot capture the diversity of behaviors observed in the archaeological record. What is needed are flexible, context-sensitive strategies capable of accommodating differences without neglecting analytical rigor. As Pomeroy and colleagues rightly observe:

> it is misguided to look for "modern human behavior" in Neanderthals, or indeed in earlier representatives of our own taxon. The use of rigid criteria based on more recent modern human analogies to identify burial or other mortuary activity is likely unhelpful, as it fails to allow for potential differences in the ways in which hominins expressed mortuary behavior. We should definitely not be forcing any expectations of a "progressive" typology ranging from mortuary to funerary behavior on to what they did. It is better to examine what Neanderthals and other hominins did, where and when, with the utmost rigor and with as few preconceptions as possible, and to try to identify what factors stimulated particular behaviors (Pomeroy et al., 2020b, p. 274).

The *variability-of-expression challenge* thus reminds us that any typological framework taking *Homo sapiens* as the implicit standard is epistemically risky if applied rigidly and without sensitivity to context. Acknowledging variability does not mean indiscriminately attributing

graves, and because of this we dismiss some deposits that lack such features, then we risk contradicting the very point raised by the variability-of-expression challenge.

complex cognition to all species, but a necessary step toward a more nuanced and comparative understanding of hominin mortuary behavior.

CHAPTER III
ADVANCING PALEOANTHROPOLOGY BEYOND DEFAULT NULLS?

1. Introduction[37]

Among the multitude of topics addressed by paleoanthropologists and archaeologists, the cognitive and behavioral capacities of our ancestors are a frequent and ambitious target. This question is the focus of cognitive archaeology, an interdisciplinary research program that aims to infer the cognitive capacities of our ancestors primarily from the material traces they left behind, though not exclusively (see the supplementary information for a brief introduction to this discipline). There have been numerous debates comparing the cognitive capacities of our ancestors with those of modern humans—particularly in the case of more recent species such as *Homo neanderthalensis* (see Villa & Roebroeks, 2014; Wynn et al., 2016)—or with those of great apes, especially when discussing the earliest stages of hominin evolution and the first makers of stone tools (see, for example, Wynn & McGrew, 1989; Wynn et al., 2011; McGrew et al., 2019). In the course of these debates, scientists sometimes refer to the concept of *null hypothesis*, a term borrowed from statistics and empirical research, which usually reflects a default assumption that there is no relationship or effect between the variables being studied. But what role do null hypotheses play in cognitive archaeology, and to what extent is their use epistemically warranted? Answering these questions is the goal of the present chapter.

The structure of the chapter is as follows: in the first part, we present some recent cases where researchers in cognitive archaeology use the term "null hypothesis". We then draw on recent debates in the philosophy of science to analyze this appeal to null hypotheses, showing that behind the use of the same expression there may lie different epistemic purposes. Next, we offer further reasons for skepticism regarding the adoption of this methodology in cognitive archaeology, and we conclude with some general reflections on possible alternative directions for future research.

As a final note, this chapter will refer to several concepts related to paleoanthropology and cognitive archaeology with which philosophers may not be familiar. To preserve the flow of the

[37] This chapter is inspired by a critical commentary that Colin Allen and I wrote for a target article for *Behavioral and Brain Sciences* (Stibbard-Hawkes, 2025). In our commentary, we partly misrepresented Stibbard-Hawkes' position, and in this chapter I will attempt to present his views more accurately, and then provide new counterarguments. Fortuitously, the title of our commentary ("*Advancing paleoanthropology beyond default nulls*") It is more appropriate for these new arguments than those in the old commentary.

I would like to thank Colin for inviting me to write the commentary with him in the first place and for his comments on this chapter. I would also like to thank the members of the Ramsey Lab, in particular Andra Meneganzin, Grant Ramsey, Gianmaria Dani, and Alejandro Fábregas-Tejeda, for their comments on an early draft of this chapter.

chapter, explanations of some of these concepts have been included in the supplementary information at the end of this chapter. In the main body of the text, we will refer to the supplementary information whenever it is useful.

2. Primitive vs Modern nulls

As noted in the introduction, this chapter focuses on the use of null hypotheses in cognitive archaeology (see the supplementary information of this chapter for a brief introduction to the discipline). In particular, we will examine two recent cases in the literature where this concept has been applied: the first concerning the cognitive abilities of members of the genus *Homo*, and the second addressing their methods of social learning. Since these will serve as our case studies and provide the starting point for the discussion, we will devote substantial space to them. Let us begin with the first case.

Stibbard-Hawkes (2025) has recently analyzed the toolkit of three extant African foragers cultures (the Hadza, the Mbuti and the G//ana) and tried to assess their "taphonomic signature", that is, the probability that their artifacts would leave enduring evidence in the archaeological record. The analysis surprisingly revealed that their symbolic culture (apart from the traded objects made of plastic and metals) is made of organic perishable materials and will not leave a clear signal in the record. In all three toolkits there is a clear bias: subsistence tools have a much stronger taphonomic signature than symbolic objects. Therefore, an archeologist of the future, millennia from now, excavating the areas where the three African foragers are living in the present would find almost no evidence of symbolic culture, and hence of modern thinking (according to the majority of archaeologists, symbolic cognition is a fundamental part of behavioral modernity, see the supplementary information). But the conclusion that for this reason these populations were not behaviorally or cognitively modern would be clearly mistaken. In addition, a second analysis also found that inter-population variability in material choice (and hence taphonomic signature) is influenced by several factors other than cognitive variation (since all three populations clearly have the same cognitive capacities), in particular ecology, mobility, and cultural evolution.

Stibbard-Hawkes draws different conclusions from his analysis, with far-reaching consequences for the debate about behavioral modernity and hominin cognitive capacities. First, his results invite a cautious stance: if his analysis is correct, we may never be able to reconstruct the details of the earliest symbolic behaviors among hominins[38]. This is because, if the African

[38] See Meneganzin & Currie, 2022, for a characterization of what they call "Rubicon expectations", the idea that there should be a clear boundary in the archaeological record that hominins cross when they become behaviorally modern.

foragers' case is projectable in the deep past, it is possible that the first expressions of symbolic culture in our ancestors are lost forever. For example, if, for whatever reason, *Homo erectus* or *Homo heidelbergensis* had a symbolic culture made only or mainly of perishable materials, we may never find a trace of symbolic cognition in these *taxa* (but see Joordens et al., 2015, for some scant evidence of symbolic, or proto-symbolic behavior in *H. erectus*). Therefore, the absence of evidence of symbolic culture in extinct hominins cannot be considered evidence of the absence of symbolic culture (see also Wallach, 2019; Thomas & Darvill, 2022; for an analysis of reasoning from absence of evidence in archaeology). This underwrites a pessimistic conclusion: our assessment of our ancestors' cognitive abilities will probably be biased towards the underestimation of their capacities, at least regarding those symbolic behaviors that have not left a trace.

To assess the soundness of Stibbard-Hawkes' analysis is beyond the scope of this chapter, but we should still bear in mind that symbolic behavior is usually considered only as one of the markers of behavioral modernity, to which ecological, technological, and socioeconomic traits are usually juxtaposed (see for example McBrearty & Brooks, 2000; see Sterelny & Hiscock, 2025 for a critical take on the view that material symbol use is an index of sophisticated cognitive capacities at all). Moreover, Stibbard-Hawkes' remark that all early symbolic artifacts could have been made of organic materials and therefore lost is a very strong assumption, since ethnography shows that modern hunter-gatherer groups also use hard materials for their symbolic artifacts (shell, ivory, bone, stone and more). However, what is of particular interest here is another conclusion he draws from his survey.

In the last two paragraphs of his target article, Stibbard-Hawkes states that, given his analyses and the fact that «the net of cognitive sophistication has almost unfailingly broadened» (p. 18), we should switch from a primitive null to a "*derived*" or "*cognitively modern*" null that assigns every species of the genus *Homo* modern cognitive capacities «until proven otherwise» and «interpret differences or shifts in material culture between past *Homo sapiens* populations … and indeed members of our genus, as resulting from extra-genetic processes» (Ivi, p. 18), that is, processes like cultural evolution, environmental variation and demographic patterns such as population size, density and migration. This last observation by Stibbard-Hawkes is in part consistent with the last two decades of research on behavioral modernity, where the appearance of "modern" behaviors in the archaeological record is no longer associated with an upgrade in the cognitive abilities of our *sapiens* ancestors, but rather with a change in our "extra-genetic" resources (see Sterelny, 2011; Scerri & Will, 2023). The novelty of Stibbard-Hawkes' proposal lies in extending this consideration to our entire *genus*.

Practically speaking, Stibbard-Hawkes' proposal implies that we should raise the inferential bar to be cleared before attributing to the members of our genus cognitive capacities inferior to ours. In other words, rather than assuming that earlier hominins had inferior abilities by default, we would need to demonstrate they truly lacked capacities comparable to our own. This means inverting the bias that researchers have typically adopted. According to Stibbard-Hawkes, this bias has long shaped, and continues to shape, the way scholars interpret the cognitive abilities of other hominins, often underestimating their potential sophistication. This proposal is especially significant for Stibbard-Hawkes, whose reply to the commentaries on his target article, aptly titled *"Hominin cognition: The null hypothesis"*, devotes considerable space to defending the rationale for adopting a cognitively modern null. We will address later the legitimacy of this move. Before doing so, we will consider another case study where the notion of "null hypothesis" pertaining to hominin cognition is invoked.

3. Hominins and early modes of cultural transmission

Tomasello and colleagues (e.g. Tomasello 1990; 1996; Tomasello & Call, 1997) famously argued that what sets human culture apart from other animals is its cumulative nature, made possible by imitation, a form of high-fidelity social learning. Unlike chimpanzees and other animals, who simply emulate end results, humans copy the process necessary to achieve a particular goal, and this is what would allow innovations to be preserved and built upon (the "ratchet effect", Tomasello, 1999)[39]. Tennie et al. (2009) build on this by introducing the concept of *Zone of Latent Solutions* (ZLS): behaviors that individuals can independently invent without the need for social learning. Most animal "traditions", they argue, fall within the ZLS, while many uniquely human behaviors lie outside it and require imitation to arise.

Tennie et al. (2016) suggest that early hominin tools (Lomekwian[40], Oldowan and Acheulean) also fall within the ZLS, meaning they could be individually invented without high-fidelity transmission such as imitation and teaching (see also Corbey et al., 2016). They propose the "island test" as a thought experiment: if a naïve hominin stranded on a desert island could reinvent the tools alone (with the right raw materials and motivations), then those behaviors lie within the ZLS. While practically unfeasible, this thought experiment can be used heuristically to derive

[39] Other presumed peculiar human features such as normativity and teaching would supplement this capacity, rendering more robust the transmission of information across generations, according to Tennie and colleagues.

[40] It has recently been stated that the archaeological finds of Lomekwi 3 (dated to 3.3 Mya; Harmanad et al., 2015) constitute the first lithic industry in history (but see Domínguez-Rodrigo & Alcalá, 2016, for a skeptical evaluation of the site dating).

archaeological predictions. Specifically, Tennie et al. argue that lithic industries falling within the ZLS should exhibit:

- *Little geographic variation*, unlike culturally transmitted behaviors, which tend to vary across space;
- *Slow temporal change*, unless affected by cultural biases or demographic patterns[41];
- *Frequent reappearances of old forms*, suggesting individual reinvention rather than cultural continuity.

Since early lithic industries show these features, this supports the hypothesis that they are independent from high-fidelity cultural transmission and instead lie within early hominins' ZLS.

This has important consequences: in the paper "*Early Stone Tools and Cultural Transmission: Resetting the Null Hypothesis*", Tennie and colleagues (2017) argue that rather than assuming cultural transmission by default, as many archaeologists seem to do, we should start with the *null hypothesis* that early stone tools are latent solutions. That is, only if archaeological data actively falsify this null (by showing patterns incompatible with individual invention plus low-fidelity social learning) we should infer high fidelity cultural transmission. Interestingly, Tennie and colleagues not only claim that Oldowan and Acheulean fall within the ZLS, but they seem to suggest that this could also apply the Middle Paleolithic stone tool industry (associated with *Homo neanderthalensis*; p. 654). The first signs of high-fidelity cultural learning, on the other hand, would date back to the Upper Paleolithic and are therefore a prerogative of *Homo sapiens*: «comparing hominin technology from the last 50,000 years both to Early Stone Age technology and to tools chimpanzees make and use today suggests that something changed in hominins between the Early Stone Age and the Upper Paleolithic (at the very latest)» (p. 654). We should thus assume that early stone tools were *latent solutions*, made without the need for teaching or imitation.

Taken together, these case studies allow us to clearly identify two distinct ways in which the notion of a null hypothesis has been applied within cognitive archaeology, which we can now briefly summarize before turning to a more precise examination:

- *Cognitively modern/derived null*: «until proven otherwise, all members of at least our genus had comparable capacities» (Stibbard-Hawkes, 2025; p. 18);

[41] However, we should keep in mind that the absence of geographical and temporal variation could also reflect the fact that these tools were already optimal for the needs of the hominins who produced them. See Sterelny and Hiscock, 2025 for a critical discussion of Tennie and colleagues' arguments.

- *No high-fidelity cultural transmission null*: early stone tools industries (maybe up until the Middle Paleolithic) did not require high-fidelity modes of cultural transmission (read imitation) (Tennie et al., 2017).

In the next section, we will attempt to provide a more precise characterization of these null hypotheses. Coming from an experimental and statistical context, the adoption of this practice is not necessarily legitimate in cognitive archaeology.

4. Null hypotheses, pseudo-nulls and heuristic nulls

The following section provides a brief introduction to the concept of null hypothesis and null hypothesis significance testing (NHST). After that, we draw on recent work in philosophy of science to assess the use of null hypotheses in cognitive archaeology.

In scientific research, the null hypothesis (H_0) is a methodological and epistemological tool: a precise baseline assumption (typically "no effect" or "no difference" between the dependent and independent variables) against which data are tested. The purpose is not to establish the truth of H_0, but to assess whether evidence justifies rejecting the null hypothesis in favor of an alternative (H_1). Two errors are possible: rejecting a true null (Type I, false positive) or failing to reject a false null (Type II, false negative). This framework was shaped by Fisher (1935) and Neyman & Pearson (1933). Crucially, the null is not a belief, but a testable proposition designed to promote caution by ruling out chance, bias, or other confounding factors.

The idea of a null hypothesis was first illustrated by Ronald Fisher (1935) through what later became a classic example in statistics: the *lady tasting tea*. A woman claimed she could tell whether milk or tea had been poured first into a cup, purely by tasting it. To evaluate this claim, Fisher devised a simple experiment. She would be presented with eight cups of tea—four prepared with milk first, and four with tea first—in random order. Her task was to identify which was which. Assuming she had no real ability and was merely guessing, Fisher formulated the null hypothesis accordingly:

H_0: *The probability of correctly identifying each cup is equal to that expected by pure chance.*

Even though Fisher did not formulate an alternative hypothesis in his original example, it is fairly straightforward to create one:

H_1: *The probability of correctly identifying each cup is greater than that expected by pure chance (hence the lady might have some incredibly sharp tasting capabilities)*

Fisher also said that the experimenter is free to choose the null hypothesis they desire, as long they respect some criteria. A quotation from Fisher himself is particularly telling: «...the null hypothesis must be exact, that is free from vagueness and ambiguity, because it must supply the basis of the "problem of distribution" of which the test of significance is the solution» (1935).

Because of its success in the experimental sciences, the term *"null hypothesis"* has often been exported into other domains. Bausman and Halina (2018) examine this trend and argue that, in many fields, certain hypotheses are granted unwarranted epistemic privilege merely by being designated as the "null hypothesis". In primate comparative psychology, for example, the hypothesis that primates predict others' actions merely by recognizing behavioral regularities, without inferring internal mental states (behavior-reading hypothesis) is often treated as the null. On the other hand, the hypothesis that they attribute beliefs, intentions, or desires (mindreading) is presented as a more complex alternative that must carry the burden of proof (e.g. Penn & Povinelli, 2007). This rhetorical move mimics the logic of statistical testing but is misleading: these are "pseudo-nulls", *full-fledged explanatory hypotheses* granted default acceptance under the guise of simplicity or parsimony, rather than null hypotheses designed to rule out chance.

As Bausman and Halina stress, the issue is that in statistics the null is usually not a substantive claim about the world but a model of background noise, and retaining it is interpreted as a negative result, not as support for the null itself. Indeed Fisher himself emphasized that the null is «never proved or established, but is possibly disproved» (1935). By contrast, pseudo-nulls encourage the fallacy of acceptance: treating failure to reject them as evidence in their favor[42]. For this reason, Bausman and Halina call for epistemic parity: competing hypotheses should be assessed by the same standards, and appeals to simplicity must be explicitly justified rather than smuggled in through the language of nulls. Recent work in philosophy of science has shown how such pseudo-nulls operate, for instance in demographic explanations of Neanderthal extinction (Currie & Meneganzin, 2022).

Now that we have differentiated between null hypotheses and pseudo-nulls, we can try to apply this distinction to our two case studies. Both Stibbard-Hawkes and Tennie et al. use the term "null hypothesis", however, as we shall see, the epistemic status of the two hypotheses is not exactly the same. While Tennie et al. consider the "no cultural transmission" null hypothesis to be a privileged one that must be empirically rejected before considering alternative hypotheses, the situation is somewhat more complex in the case of Stibbard-Hawkes. Although he repeatedly states

[42] For a more detailed argument against the use of pseudo-nulls, see the paper by Bausman and Halina. For examples of how, in some cases, treating certain explanations as null hypotheses is a way of penalizing them rather than privileging them, see Lloyd, (2015).

in his response to the commentary that his is not a statement about a state of affairs in the world, it should also be noted that he himself is ambiguous in presenting his null hypothesis, saying that adopting it would entail that «until proven otherwise, all members of at least our genus had comparable capacities». This formulation brings its null very close to the pseudo-nulls identified by Bausman and Halina, and seems to indicate that it is not only a "comparator" necessary to temper our *sapiens*-centric biases, but also a belief to be maintained until it is refuted.

This is how initially, in the commentary written for Stibbard-Hawkes (2025), we interpreted the cognitively modern null: as a statement of belief or conclusion of SH's argument, a change in the default assumption regarding our genus: given the results of the survey, and given that much of our symbolic material culture may have been lost, we should grant all members of our genus the same cognitive abilities (Bedetti & Allen, 2025, see also Irurtzun's commentary for a similar interpretation). Later, in his response to the commentary, Stibbard-Hawkes clarified his position, arguing that his null hypothesis should be understood in the statistical sense of the term, not as a hypothesis to be refuted before demonstrating that other *Homo* species had abilities different from ours. As already mentioned, the function of the new null hypothesis should be to raise the standards required to demonstrate lower cognitive complexity in other hominins, and temper our biases:

> Like both Bedetti & Allen and Irurtzun, I find it most parsimonious that earlier members of both our lineage and our genus did not have brains, minds and cognitive capacities identical to ours. Adopting a "cognitively modern" null is wholly compatible with this belief. The new null model simply changes the inferential bar we must clear. As Vasil highlights, it prevents the more parsimonious primitive/ancestral null from being tacitly accepted, and forces us to think clearly about how differences would manifest using a probabilistic "evidence-based approach" (p. 57)

> In failing to overturn it, and in reflecting upon that failure, we might moderate our beliefs somewhat. If we are ever to overturn it beyond reasonable doubt, we must leverage all available evidence. (p. 57).

So, while Tennie et al's null is, to all intents and purposes, an explanatory hypothesis, Stibbard-Hawkes conceives of his as a genuine null hypothesis in the strict sense, a "comparator" that defines evidential thresholds, not a factual assertion about the world susceptible to proof. The requirement for simplicity also distinguishes the two null hypotheses.

> This simple shift in perspective prioritizes the systematic investigation of more *parsimonious* potential explanations and forces us to demonstrate, rather than presume, that stone tools could not have existed without high-fidelity cultural transmission. (Tennie et al., 2017, abstract)

It is easy to understand why no high-fidelity cultural transmission is a more parsimonious hypothesis. Tennie and colleagues consider imitation to be a cognitively sophisticated way of learning and a prerogative of human beings. For this reason, assuming that our ancestors used

simpler and more phylogenetically widespread methods of cultural transmission would be the most parsimonious hypothesis, at least for simpler stone tool industries. The logic mirrors that found in the contrast between the behavior-reading and mindreading hypotheses. Stibbard-Hawkes' case is a little trickier:

> In light of present findings, it appears more *parsimonious*, at least in the absence of other conclusive evidence, to interpret differences or shifts in material culture between past *H. sapiens* populations (…) and indeed members of our genus, as resulting from extragenetic processes. (Stibbard-Hawkes, 2025, p. 18).

But later he seems to contradict himself:

> I find it most *parsimonious* that earlier members of both our lineage and our genus did not have brains, minds and cognitive capacities identical to ours (p. 57, emphasis added).

So, is the derived null more parsimonious or not? It can be argued that Stibbard-Hawkes contradicts himself precisely because he shifts back and forth between two conceptions of his derived null, at times as a null hypothesis, at others as a substantive alternative[43]. Nevertheless, his remarks can be interpreted in a way that reconciles this tension. At first glance, assigning all *taxa* of our genus a level of cognitive sophistication comparable to our own seems the opposite of the most parsimonious hypothesis. But what Stibbard-Hawkes seems to mean by "simpler" here is something like "in accordance with the rest of our knowledge about human evolution," and the fact that it introduces the new null as more parsimonious "in light of present findings" (p. 18) is telling[44]. This would explain why later in his response to the commentaries, Stibbard-Hawkes claims the opposite: the cognitively modern null would then be more complex in the usual sense of involving more or more complex entities, causes, or processes, but simpler in the sense of being in accordance with our background knowledge of human evolution, in light of his findings[45].

Thus, we will grant Stibbard-Hawkes that his derived null functions as a *comparator* rather than as a substantive claim about the world (i.e., a pseudo-null). Nevertheless, his proposal still involves a degree of conceptual confusion, which gives rise to the contradictions highlighted above. Nonetheless, it is clear that it is not a null hypothesis in the strict sense. We are not in an experimental context, and the derived null is not simply a model of background noise that can confound the causal relationship between two variables. Stibbard-Hawkes' null is neither a null

[43] A further cause of confusion is the fact that the opposite perspective, which Stibbard-Hawkes calls "primitive" or "ancestral" null, is in fact a default assumption about the cognitive abilities of our ancestors (see pp. 5-6 for example).

[44] This analysis aligns with John Norton's material theory of induction: there is no direct connection between simplicity and likelihood; in reality, appeals to simplicity are justified on the basis of inductive inferences grounded in the background knowledge of the relevant discipline(s) (e.g. Norton, 2021).

[45] In what sense a comparator could be more or less parsimonious or simple is a question we leave open. It can be argued that, once again, the issue arises from Stibbard-Hawkes's conflation of the two meanings of *null hypothesis*.

hypothesis nor a pseudo-null, but something in between the two. We can call these types of hypotheses "heuristic null", to stress the fact that they are not explanatory hypothesis but methodological baseline adopted to calibrate evidential standards. But even granting the most charitable interpretation as a heuristic null, his strategy is not without epistemic problems.

5. General objections

Tennie et al.'s *no cultural transmission* null exhibits all the features of a pseudo-null hypothesis: it is treated as the default position that must be rejected before alternative explanations can be accepted, and failure to do so is taken as evidence that it offers the best account of the phenomenon under study. Moreover, its epistemic privilege derives from its presumed parsimony. Bausman and Halina (2018) have already provided several arguments against the use of pseudo-nulls, and we refer to their paper for further critical discussions of this strategy (see also Smith, 2018; Currie & Meneganzin, 2022). In this section, we will focus on Stibbard-Hawkes' cognitively modern heuristic null, and explain why this strategy is not warranted in cognitive archaeology. In the next section we will cover what could be the most profitable path to take.

Stibbard-Hawkes acknowledges that he is employing null hypotheses in a context that differs from their original domain of application: «even in Bayesian or informal inferential reasoning, null models still have utility as comparators» (p. 57). This is precisely where the problem lies: the use of null hypotheses may seem useful, but in reality, risks being counterproductive. We would like to recall how Ronald Fisher, one of the fathers of null hypothesis significance testing, said that the null hypothesis can be chosen freely, provided that it is not ambiguous or vague and its conditions for refutation are clear (1935). The Lady Tasting Tea experiment exemplifies this condition: the null hypothesis has extremely precise conditions for refutation, namely, that the lady should be able to identify whether milk or tea was poured first into the cup with a probability greater than expected by chance. However, the same cannot be said in the case of the derived null.

The first critical issue, and perhaps the most serious, concerns the *testability* of the cognitively modern null. What type of evidence could disprove it? Archaeological findings *per se* clearly cannot. On the one hand, finding evidence of symbolic behavior in earlier hominin species would count in favor of the derived null. On the other hand, the persistent lack of evidence of symbolic artifacts for other species is perfectly compatible with the idea that, if they existed, they were made of perishable materials, and therefore there is no way of refuting the derived null by material findings alone (a point raised also in Irurtzum's commentary). What about other types of evidence? To be fair, Stibbard-Hawkes provides some guidance on which type of non-archaeological data

could indirectly disprove his null. In a paragraph aptly called "*Beyond reasonable doubt: Overturning the derived null requires extraordinary evidence*" (pp. 57-58), he engages with ours and others' commentaries that highlighted this issue of testability (again, see especially Irurtzun, 2025). In the same paragraph, Stibbard-Hawkes suggests some strands of evidence that might, in the future, refute the derived null: correlation between brain size and cognitive capacities, clues about the relative expansion of different brain regions, anatomical features related to the evolution of language (e.g. laryngeal and ear morphology) and ancient DNA. After providing a brief overview of these types of evidence, Stibbard-Hawkes concludes that none of them can currently overturn his null hypothesis. Colin Allen and I proposed in our commentary a (weak) correlation between relative brain size and cognitive capacities as (current) positive evidence against the derived null (2025). Although the archaeological record does not reveal a strict one-to-one correspondence between relative brain size and cognitive abilities, the overall pattern suggests that increases in the former are generally accompanied by expansions in the latter—as reflected in broader behavioral repertoires, more versatile toolkits, the colonization of new environments and so on. It is difficult to overlook how, from *Homo ergaster* (with a cranial capacity of about 800 cc), through *Homo erectus* (≈1000 cc) and *Homo heidelbergensis* (≈1250 cc), to *Homo sapiens* and *Neanderthals* (≈1450 cc), the archaeological record indicates an overall increase in the complexity of hominin lifeways. In this regard, Stibbard-Hawkes misinterprets our claim: we are not stating that this correlation could eventually falsify the derived null, but rather that *at the present moment*, it provides reasons *against the setting* of the null in the first place.

However, the problem of testability remains: the derived null does not follow Fisher's dictum that a null hypothesis should always have clear conditions of falsifiability. This is one of the strongest arguments for why we should not import null hypotheses into non-experimental contexts. While null hypotheses have strict conditions of falsification in their original domain of application, the same strict conditions cannot be guaranteed in contexts such as that of cognitive archaeology. Stibbard-Hawkes has pointed to categories of evidence that could one day do so, but the criteria remain too vague. Appealing to ancient DNA, neuroanatomical correlates, or language-related anatomy provides only general directions, not concrete empirical predictions whose failure would count decisively against the null[46]. For instance, what specific discovery in ancient DNA would

[46] In comparison, when presenting their null hypothesis, Tennie et al. are fairly clear in listing the evidence that could refute it and that which could not. They propose possible ways to falsify their null hypothesis: by identifying lithic features inconsistent with patterns expected from individually learned behaviors. One method is to estimate how often a behavior should appear without cultural transmission and compare it with archaeological frequencies, though this is methodologically difficult. Another is to compare contemporaneous nearby assemblages: greater variability may indicate independent invention, while similarity may suggest cultural transmission. However, evidence such as technological complexity, failed modern replications, or apparent behavioral stasis cannot alone refute the null, since these can arise from individual learning or latent solutions.

count as evidence against the derived null? The path from fossil genes to the cognitive capacities of the organisms that carried them is long and uncertain, and without precise preliminary specifications, it is unclear how such evidence could ever falsify the derived null. The consequence is that it is not even clear how we should revise our beliefs. When faced with failed attempts to falsify the derived null, we should temper our *sapiens*-centric biases and revise our assumptions, according to Stibbard-Hawkes. However, if the derived null cannot practically be falsified, two outcomes are possible. We may leave our prior beliefs intact, making the null effectively pointless. Or we may simply replace one bias with another, swinging to the opposite extreme. Once a hypothesis is shielded from empirical refutation, reversing that shift becomes exceedingly difficult.

There are also elements of *arbitrariness* in the choice of the null, as we stressed in our commentary. The null is not grounded on direct (i.e., not mediated by inferences), positive evidence, in the sense that at the moment there is no indication that *taxa* other than *Homo sapiens* and Neanderthal shared comparable cognitive capacities (even in the case of *Neanderthals*, however, this remains controversial, e.g., Mithen, 2024, for a recent example). The cognitively modern null is based instead on the indirect datum that the non-traded symbolic objects of three African contemporary forager groups would not leave an enduring trace in the future archeological record. In our commentary we remarked:

> First, what is the evidence in favor of the assertion that the earliest representatives of *Homo* were cognitively modern? Why place the boundary of *Homo-sapiens*-level cognition at that precise point rather than another? It is true that absence of evidence is not evidence of absence, but this doesn't mean that every case of absence of evidence is the result of taphonomic biases or other phenomena. Sometimes, there is no evidence of a certain behavior in the fossil and archeological records because that species was not capable of the behavior (p. 25, see also the commentary by Irurtzun).

To which Stibbard-Hawkes replies:

> Our genus is an intra-taxon grouping. Although less similar than members of the same species, a certain homogeneity of traits is a core assumption for genus-level groupings also. We should not expect similarities to cease above the neck. It is easy to forget this, and the derived null provides a heuristic check on our biases (p. 58).

It's true that we shouldn't expect the similarities to stop at the neck, but this is exactly the point: the genus *Homo* shows great internal variability, both above and below the neck, and there are significant differences, both in terms of absolute and relative brain size and shape. To cite one of the possible comparisons, the differences between *Homo sapiens* and *Homo naledi* are significant from all points of view (Garvin Heather et al., 2017; Holloway Ralph et al., 2018; Hurst et al., 2024). The evolutionary history of our genus is far from typical: it is the story of a lineage

that experienced a gradual, though far from linear, increase in cranial capacity, accompanied by the diversification of the ecological and sociocognitive niches occupied by its members. Therefore, we reiterate the question: why place the null right there on that boundary and not somewhere else?

Stibbard-Hawkes would probably point out that the cognitively modern null only serves to temper our biases; and indeed this is one of the main arguments in favor of his proposal (see especially section *R3.3* of his response). Again, according to Stibbard-Hawkes, a persistent *sapiens*-centric bias has long shaped the reconstructions of our past, leading researchers to treat *Homo sapiens* as the benchmark for cognitive and cultural sophistication. As a result, material and behavioral evidence from other hominins, most notably Neanderthals, has often been interpreted through a deficit lens and dismissed (see also Fuentes et al., 2025). However, these assumptions have been repeatedly challenged by recent archaeological discoveries, particularly concerning Neanderthals, which show that other species engaged in sophisticated and symbolic behavior, although with relevant differences (Wynn et al., 2016; see also supplementary information). According to Stibbard-Hawkes, the adoption of the derived null would encourage the attribution of complex cultural practices to other species, promoting a sort of reparation for *sapiens*-centric bias. For instance, the perforated bear femur from Divje Babe I Cave (see Figure 3) would likely have been immediately recognized as a Neanderthal-made flute had we adopted the derived null (see Turk et al., 2020, for a review of the debate).

Figure 3. Divje Babe flute, picture by Petar Milošević, https://it.wikipedia.org/wiki/File:Divje_Babe_flute_%28Late_Pleistocene_flute%29.jpg

Several considerations can be made in this regard. One might argue that in some instances the clash between two opposing camps has, in fact, benefited paleoanthropological research. For example, the debates on the perforated bear femur from Divje Babe I Cave prompted the use of

advanced analytical techniques to distinguish genuine artifacts from traces produced by carnivore activity (see again Turk et al., 2020)[47]. But an analysis of these issues would require at least another thesis. One might instead ask whether Stibbard-Hawkes' null could be reformulated to avoid the issues outlined above. Alternatively, it may be that his derived null is beyond repair, yet there could still be room for other heuristic nulls within cognitive archaeology. However, given the epistemic context of this field, the use of such heuristic nulls seems bound to generate problems of testability. Rather than pursuing that path, the following section explores how cognitive archaeology can proceed without relying on any form of null hypothesis.

6. What to do instead

Inference to the best explanation (IBE) is a form of inductive reasoning extensively analyzed in philosophy (see Harman, 1965; Lipton, 2004; Douven, 2025). It involves inferring, from the premise that a given hypothesis would provide a better explanation for the available evidence than any alternative hypothesis, to the conclusion that the given hypothesis is true, or, in the probabilistic context of archaeology, that it is more likely than competing hypotheses (Harman, 1965, p. 89). Inference to the best explanation has been regarded as the main argumentative strategy in archaeology, common even to distinct theoretical traditions such as processualism and post-processualism (Hanen & Kelly, 1989; Fogelin, 2007; Fotiadis, 2018; Campanaro, 2021).

But how can we determine whether one hypothesis is superior to another? Several criteria have been proposed in both philosophical and archaeological literature. Fogelin (2007), in his analysis of IBE in archaeology, identifies seven features of successful explanations: *empirical breadth* (the capacity to integrate multiple lines of evidence) and *generality* (the capacity to account for a wide range of phenomena), both of which must be balanced by *modesty*; *refutability* (a good explanation should be falsifiable); *conservatism* (it should not contradict well-established explanations or principles unless it passes a particularly stringent test); *simplicity* (it should rely on fewer or less complex entities and mechanisms); and *multiplicity of foils* (its ability to explain why a particular event occurred in one way rather than another). Applying these criteria to our hypotheses can guide us in evaluating which is better, though, of course, reality is far more complex than any schematic set of principles can capture.

[47] It may be argued that caution in interpreting the Divje Babe bear femur as a flute is, at least in part, justified. The find would represent the earliest evidence of musical ability in Neanderthals, thereby pushing back the emergence of such capacities within the entire *Homo* genus. By contrast, in the case of *Homo sapiens*, the presence of a musical culture is known *a priori*. Whether this reflects a *sapiens*-centric bias or rather a healthy dose of epistemic caution remains an open question.

IBE has also been identified as the main form of reasoning exemplified by inferences in cognitive archaeology (e.g., Pain, 2019; Killin & Pain, 2023), and we argue that it provides a more appropriate framework than heuristic null reasoning. Stibbard-Hawkes maintains that the derived null «forces us to think clearly about how differences would manifest using a probabilistic, evidence-based approach» (p. 57). Yet this is precisely what IBE already requires. It encourages us to consider the various hypotheses on the same level, paying particular attention to which is most empirically adequate and/or possesses the greatest number of epistemological virtues that we presented above. It encourages us to consider how the various lines of evidence converge (or not) in providing support for the same hypothesis, the weight that these individual lines have, etc. There is nothing added by null hypotheses that is not already encompassed within standard archaeological methodology (Bausman & Halina, 2018, also assert something along this line in relation to the use of pseudo-nulls).

Stibbard-Hawkes emphasizes that adopting the cognitively modern, or derived, null might help correct our anthropocentric and *sapiens*-centric biases. Yet this raises the question of why we should replace one bias with another, whether there is actually such a bias or not (see also Fitzpatrick, 2008, for the same remark on Morgan's canon in comparative cognition). Questions such as whether Neanderthals were as cognitively sophisticated as *H. sapiens*, or whether the perforated bear femur from Divje Babe I is a flute, are best addressed by asking which hypothesis provides the superior explanation of the available evidence, without introducing asymmetries in how we evaluate rival claims. Clearly, scientists will always be subject to biases that affect the weight they assign to different pieces of evidence or the way they interpret data; yet the ultimate arbiter should remain the empirical domain itself. This is a proper way of addressing these biases. This also preserves the benefits of pluralism, without having to resort to null hypotheses outside their original context of application. If other hominins prove to be cognitively inferior to us, it will be because the best of our scientific knowledge, at this moment, tells us so. When, instead, the evidence does not allow us to choose between different hypotheses, it would be more reasonable to suspend judgment and wait for science to run its course.

7. Conclusions

As stated by several philosophers (e.g. Alison & Wylie, 2016; Currie, 2018) one of the strengths of historical scientists is their capacity to integrate resources, technologies and methods from other disciplines. There is no denying how successful these efforts have been. Radiocarbon dating and ancient DNA extraction have generated new lines of evidence that were previously unthinkable and are now routinely used in archaeological and paleoanthropological excavations.

Less attention has been paid to the fact that borrowing resources from other disciplines is not always beneficial. The use of null hypotheses in cognitive archaeology is one such case. Developed in a specific statistical and experimental context, null hypotheses should not be used outside of the said background, because that leads to the criticalities identified in this chapter. Not only is the use of pseudo-nulls problematic (as diagnosed by Bausman & Halin, 2018; Currie & Meneganzin, 2022), but also the use of heuristic nulls raises several critical issues. Cognitive archaeology is not an experimental discipline, and the adoption of heuristic nulls can only lead to problems of falsifiability. Not always does the borrowing of conceptual tools pay off.

When dealing with the cognitive capacities of our ancestors, a more fruitful approach would be to treat all assumptions equally, and evaluate them by considering their empirical adequacy and eventually their theoretical virtues (simplicity, explanatory power, consistency, predictive power, and so on). What were the cognitive capacities of *Homo erectus*, or whether the Acheulean handaxes production required imitation or just emulation are empirical questions that have to be settled by comparing different hypotheses with the archaeological record, and the one that is best explained by the latter is considered (fallibly) to be the correct one.

This is not to deny that null hypothesis significance testing has legitimate applications within paleoanthropology and related disciplines. On the contrary, in the right contexts, this methodology can be used profitably. Stibbard-Hawkes himself presents a case where the null hypothesis is applied correctly and therefore is easy to test. In responding to Bedetti & Allen (2025), he cites as an example a study he co-authored on the relationship between schooling and eyesight in the Hadza (Stibbard-Hawkes & Apicella, 2022). The null hypothesis of the study was that there is no causal relationship between school and vision deterioration, while the alternative hypothesis was that such a relationship does exist. Stibbard-Hawkes and Apicella found no statistically significant correlation between the two variables in the Hadza, and therefore were unable to reject the null hypothesis. As Stibbard-Hawkes correctly pointed out, this does not mean that the null hypothesis is true (see also Smith, 2018), and the lack of correlation between worsening eyesight and school may have a multitude of different explanations. But this is a fair use of the NHST, one where the conditions of rejection of the null hypothesis are clear (but again, see Smith, 2018; 2020 for critical remarks on the use of NHST in evolutionary anthropology). When, on the other hand, we are dealing with hypotheses that are not easily translated into simple quantitative judgments, as in the case of the "cognitively modern null", the situation is epistemically more complex.

SUPPLEMENTARY INFORMATION
KEY CONCEPTS AND DEBATES

Cognitive archaeology: a primer

Reconstructing the cognitive abilities of our ancestors is one of the most ambitious projects in paleoanthropology and archaeology. Brains do not fossilize, just as many of our behaviors leave no trace in the fossil record. For these reasons, attempts to trace the evolution of our cognitive abilities must be based on very indirect evidence. This has led several researchers to voice a certain pessimism about the success of this endeavor. Geneticist Richard C. Lewontin expressed this well in a book chapter, quite tellingly named *The Evolution of Cognition: Question We Will Never Answer*:

> I must say that the best lesson our readers can learn is to give up the childish notion that everything that is interesting about nature can be understood. History, and evolution is a form of history, simply does not leave sufficient traces, especially when it is the forces that are at issue. Form and even behavior may leave fossil remains, but forces like natural selection do not. It might be interesting to know how cognition (whatever that is) arose and spread and changed, but we cannot know. Tough luck. (1998)

Despite Lewontin and others' pessimism, there are scientists who are attempting to reconstruct the evolution of human cognition. Among these efforts, cognitive archaeology has emerged in recent decades as a specialized field devoted to investigating the cognitive abilities of our ancestors. Cognitive archaeologists infer these capacities from the material remains left behind by hominins, (in the early phases of our evolution especially stone tools), mobilizing and leveraging theory, methods and evidence from a broad spectrum of disciplines. Normally, cognitive archaeologists rely on cognitive science models to infer the cognitive abilities necessary for the construction and use of a particular artifact[48]. However, cognitive archaeology methods are not limited to the application of cognitive sciences to the archaeological record but can involve other streams of evidence, such as primatology, comparative psychology, ethnography, genetics, brain imaging studies, and many more.

[48] For an accessible introduction to the discipline see Wynn & Coolidge, 2022, for philosophical introductions to inferences made in cognitive archaeology see Currie & Killin, 2019; Pain, 2021; Currie & Meneganzin, 2022; Killin & Pain, 2023. For another approach to the study of the evolution of our mind, which does not refer to archaeological evidence but to that of contemporary psychology, see Chapter 4.

Behavioral modernity and the Neanderthal mind

Another concept closely related to the study of the evolution of the human mind is that of *behavioral modernity.* Tracing the decades-long debate around what behavioral modernity is and how is recognized in the archaeological record is a daunting task, even more so given that the very concept of "behavioral modernity" has evolved over time, following new archaeological findings (see Meneganzin & Currie, 2022 for and overview of how the concept changed through time and Scerri & Will, 2023, for a recent summary of the relevant archaeological record).

Initially, the debate revolved around the apparent time lag between the appearance of anatomically modern *sapiens* (between 300 to 200 ka; Stringer, 2016) and modern behavior (usually taken to mean "falling within the range of expressions of ethnographically known foragers", e.g. Sterelny, 2011; Currie & Menganzin, 2022), which seems to have first appeared around 40 ka in Europe. This puzzling delay was named by the archaeologist Colin Renfrew the "Sapient Paradox" (Renfrew, 1996). Among the behaviors that we consider as modern there are the manufacture of composite tools, blades and microblades, bone tools, long-distance exchange of raw materials, occupation of previously uninhabited areas and symbolic behavior such as self-adornment, cave art and burials (see McBrearty & Brooks, 2000, for a comprehensive list of these traits).

At the turn of the century, the realization that an unwarranted focus on the European Upper Paleolithic record biased the research on behavioral modernity, and the finding in the Middle Stone Age African record of many instances of modern behavior caused a shift in the research agenda. Most scholars have moved from the idea that behavioral modernity was caused by a sudden leap in *sapiens* cognition thanks to a genetic mutation changing brain architecture (a view held, among others, by Mithen, 1996; Klein, 2008) to a more gradual and mosaic assemblage of the package of modern behaviors, caused most likely by demographic and sociocultural factors, instead of cognitive, intrinsic changes (Sterelny, 2011).

At the same time, there has been an "extension" of behavioral modernity to Neanderthals, overturning the popular image of them as brutish and primitive cavemen. Being our closest relative, often their cognitive abilities were compared with ours in an attempt to understand whether and eventually in what respects they were cognitively inferior to us (or different, see for example Wynn et., 2016). Until very recently, the consensus in the scientific community was that Neanderthals were cognitively inferior to us in many domains (see Peeters & Zwart, 2020; Nowell, 2023, for retrospective reviews). The Neanderthal fossil and archaeological records, lacking traces of many of the markers of a behavioral modernity, were seen as inescapably indicating a mind less sophisticated than ours.

But in recent years, thanks to new archaeological discoveries and new techniques such as ancient DNA analyses, the consensus seems to be shifting in the opposite direction (see for exa,mple Sykes, 2020 for a review). Notable findings include the use of perforated and ochre-stained shell beads, interpreted as personal ornaments (e.g. Zilhão et al., 2010); the manufacture of composite instruments (Niekus et al., 2019), the exploitation of birds of prey feathers and claws for aesthetic purposes (Peresani et al., 2011), and cave art in parts of Spain that has been dated to periods long before modern humans arrived (Hoffmann et al. 2018; but see White et al., 2020 for criticisms), and many more. The discovery that Neanderthals engaged in many of the behaviors that we deem as modern is prompting a reconsideration of their cognitive capacities, seen as increasingly comparable to our own. The paleoartistic representation of Neanderthals is changing as well. From the old *topos* of the Neanderthal as a primitive, brutish caveman, in the light of new evidence we have moved on to images of Neanderthal individuals with bodies adorned with bird feathers, necklaces made of shells and dyes of red ochre. Nevertheless, there are still researchers who are skeptical about the cognitive abilities of Neanderthals, see for example Mellars 2010; Mithen, 2014; 2024.

Homo naledi, again

As discussed throughout this chapter and in the supplementary materials, considerable attention has been devoted to reconstructing the cognitive capacities of other hominins. However, views on this matter diverge markedly (see, for instance, the contrast between the primitive and derived null hypotheses discussed by Stibbard-Hawkes). Unsurprisingly, the debate surrounding *Homo naledi* likewise reflects this division into two opposing camps. In responses to the 2023 preprints (Berger et al., 2023a), some reviewers argued that the starting null should be natural deposition: bodies were naturally covered and only "extraordinary" and *multiple* lines of evidence could justify intentional burial (see also Martinón-Torres et al., 2023; Val, 2016). This position rests on what Stibbard-Hawkes (2025) calls a *primitive* or *ancestral* null: given *naledi*'s small endocranial volume, the default expectation should be the absence of "meaning-making" behaviors (burials, engravings).

The *naledi* team contest this assumption. In their response to the reviews, they acknowledge having initially adopted a "conservative view", i.e. that small-brained hominins were unlikely to engage in symbolic practices and that Rising Star deposits were natural. This had a tangible impact on their research, leading them to favor the natural accumulation hypothesis and thus overlooking evidence that what they were studying may have been burials (Berger et al., 2023a).

The third preprint (Fuentes et al., 2023; see Fuentes et al., 2025 for the version of record) explores the evolutionary implications of the other two: if intentional burial and rock engravings are enacted by a small-brained hominin, then the oft-presumed coupling between brain size/EQ and symbolic behavior requires revision. Meaning-making practices may have emerged within a sociocultural niche characterized by coordination and cooperation, despite the absence of any significant increase in brain size (Kissel & Fuentes, 2018, 2021). Such a reframing would also affect how we treat other small-brained hominins: for instance, reassessing earlier skepticism that associated *H. floresiensis* with tools deemed "too complex" given its cranial capacity (Lahr & Foley, 2004; Brumm et al., 2006).

As we said, this clash between two opposing camps is recurrent in paleoanthropology and archaeology: on one side those who hold a conservative stance, and on the other those who are willing to attribute to extinct *Homo* species cognitive abilities comparable to ours. This kind of polarization recalls an older and broader discussion about the mental faculties of non-human animals, long debated within comparative cognition.

Romantics vs killjoys: parallels in comparative cognition

Since the early formulation of Morgan's canon (Morgan, 1903), researchers in comparative psychology have argued about the correct explanations of animals' behaviors. The core issue is whether we should privilege simple associative mechanisms or whether we should resort to the attribution of more complex cognitive mechanisms such as reasoning, planning, or mindreading. Morgan's principle famously stated that «in no case is an animal activity to be interpreted in terms of higher psychological processes, if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development» (1903, p. 292). The well-known case of Clever Hans, the Berlin horse once believed to solve arithmetic problems but later shown by Pfungst to respond to subtle human cues, has often been used to illustrate this principle. The episode revealed that apparently complex performances can be explained by associative mechanisms, without invoking higher-order reasoning, exactly as Morgan advised (but see Fitzpatrick 2008 for a different reading).

Discussions about Morgan's canon have continued for over a century (de Waal 1999; Sober 2005; Fitzpatrick 2008; Andrews & Huss 2014; Meketa 2014). Despite the disagreements, the consensus is that it has generally been used to justify the preference for the attribution of "lower" cognitive mechanisms to non-human animals (Allen 2006; Shettleworth 2010b). This preference is often justified on the basis of parsimonious reasoning: associative processes are seen as simpler

than cognitive ones, so they should be epistemically privileged whenever they are available as an explanation of the performance of an animal. Only when associative accounts are ruled out it becomes reasonable to appeal to more complex mental processes.

Similarly, Dennett (1983) placed hypotheses in animal cognition research on a continuum going from "romantic" to "killjoy", where romantic hypotheses ascribe high cognitive sophistication to non-human animals, while killjoys ascribe low cognitive sophistication, and he took "killjoy" to be the default null. Others have argued against this default (e.g. Sober, 2005; Andrews & Huss 2014; Mikhalevich 2015). These attitudes affect the kinds of experiments are pursued in comparative cognition and the cognitive abilities attributed by scientists to the animals they study, but first, affect the choice of the null hypothesis about the cognitive capacities of the animal under study. Killjoy researchers typically privilege associationist hypotheses, while romantics prefer high-level cognitive hypotheses. In the long-standing debate on primates' theory of mind, for instance, the former position is exemplified by Povinelli and colleagues, who interpret chimpanzees' performance in experimental tasks in terms of associationist behavior-reading (e.g., Penn & Povinelli, 2007); while romantics are represented by Tomasello and colleagues who instead claim that great apes are to some extent capable of mindreading (e.g. Tomasello, 2023).

A similar pattern is recurrent in debates within paleoanthropology and cognitive archaeology, as we showed. But other parallels can be found between cognitive archaeology and comparative cognition. This should not be surprising: hominins are animals, and studying their cognitive capacities is, in essence, another case of comparative cognition, though one that deals with long-extinct species.

Killin and Pain (2023) recently pointed out that Thomas Wynn, one of the founders of cognitive archaeology, often repeated what Overmann and Coolidge (2019, p. 3) called a "mantra": make no more assumptions than necessary to explain a phenomenon. This idea first appeared in discussions about the Oldowan lithic industry. According to Wynn and McGrew (1989, p. 384), the archaeological record can only reveal the minimum competence required to produce specific tools, that is, the minimal cognitive capacities consistent with the evidence. Killin and Pain argue that this reasoning appeals to parsimony: we should prefer the simplest explanation that fits the data, involving as few unsupported assumptions as possible. Thus, if great apes can make similar tools, and no evidence points to more complex technologies, the Oldowan toolmakers should not be credited with greater intelligence than that.

Wynn's view strongly recalls Morgan's canon. He writes that «the rule of parsimony must be applied... favoring the simplest form of cognition that can account for the evidence» (Wynn 2016,

p. 208), and «the simplest cognitive system that can account for archaeological features must be given priority» (Coolidge et al. 2016, p. 221). In this sense, his "mantra" serves in cognitive archaeology much the same role that Morgan's canon plays in comparative cognition, offering a methodological foundation for more conservative interpretations. Stibbard-Hawkes and the *naledi* team, by contrast, resemble those who have questioned the canon's authority. The disagreement becomes explicit when we recall Wynn's claim that «sometimes absence of evidence is a very persuasive argument for absence» (Wynn 2000, p. 114). Stibbard-Hawkes's response is the opposite: the lack of material traces should not be taken as proof that complex cognition was absent.

We have seen some similarities between comparative cognition and cognitive archaeology: the division into two camps with opposing views, the presence of similar methodological principles (and of similar epistemological problems), as well as a clear interest in investigating the cognitive abilities of other animals. But if the two disciplines address similar problems, then perhaps the solutions adopted by one may have their own utility for the other as well; a connection that, to date, has received little attention from philosophers (but see Bedetti & Allen, 2025, for some suggestions)[49].

[49] There may be some inconsistencies with the final paragraphs of the chapter, but these are only apparent. While here we have described how debates on human evolution have historically unfolded, in the chapter's conclusions we wanted to point out what could be a better alternative, where the focus is not on biases and opposing sides.

CHAPTER IV
ONCE UPON A TIME IN THE PLEISTOCENE: INTEGRATING PALEOANTHROPOLOGY AND EVOLUTIONARY PSYCHOLOGY[50]

1. Introduction

In this final chapter we will look at the prospects of integrating findings from paleoanthropology and archaeology (and by extension other human evolutionary sciences) with other disciplines. How do other fields use, or how should they use, the evidence and theoretical resources provided by paleoanthropology? More often than not, the situation is the reverse. The historical sciences are often praised for their remarkable ability to draw upon and integrate methods and insights from other fields (see, for example, Currie, 2018). One well-known example is the adoption in archaeology of ^{14}C dating, a method originally developed by physicists (see Chapman & Wylie, 2016, chap. 4, for a reconstruction of the long and complex epistemic process that led to the fine tuning of this technique for the archaeological context). Another striking example is the more recent, and extraordinarily successful, incorporation of ancient DNA analysis into paleoanthropology (see Downes, 2019, for a philosophical reflection on the significance of this development). It is difficult to exaggerate the role that these borrowed techniques and bodies of knowledge can play.

While this phenomenon certainly deserves careful philosophical reflection, in this final chapter we will argue that attention should also be directed toward the opposite flow of information: from paleoanthropology and archaeology to other disciplines. Although paleoanthropology has traditionally borrowed more than it has lent in terms of technologies and methods, evidence concerning human evolution nonetheless plays a pivotal role in several neighboring fields. Indeed, research programs that seek to reconstruct our psychological and behavioral evolution depend (also) on data provided by paleoanthropologists and archaeologists. Human behavioral ecologists, evolutionary psychologists, cultural evolutionary theorists, and related scholars must rely on fossils, artifacts, ancient DNA, and the inferences drawn from these sources as the fundamental epistemic scaffolding for their hypotheses (for an introduction to these

[50] The inspiration for this chapter stems from my visiting period at the Max Planck Institute for Geoanthropology in Jena, where the relationship between humans and their environments is a central area of study. Engaging with researchers focused on paleoenvironmental reconstruction—particularly on the role tropical forests have played in human evolution—has greatly shaped the ideas developed here. In this regard, I would like to express my special thanks to Noel Amano, Nicolas Bourgon, Victor L. Caetano Andrade, Deepak Jha, Rachel Rudd, Patrick Roberts and Michael J. Ziegler.

The content of this chapter has been presented in condensed form in my book *La mente preistorica: analisi epistemologicala della psicologia evoluzionistica.*

disciplines see Brown & Lala, 2024). For instance, human behavioral ecologists seeking to formulate adaptive hypotheses concerning the evolution of childhood cannot disregard the paleoanthropological record, which offers crucial evidence regarding when this life stage emerged, in which ecological and social contexts, and so on. To neglect such data would amount to speculation detached from historical reality (see Meneganzin & Currie, 2025; for a recent introduction on investigating development through paleoanthropological data see Lequin et al., 2025).

In this chapter, we focus on one of the sciences concerned with the evolution of human behavior, *evolutionary psychology*. The aim is to discuss how evolutionary psychologists integrate (or should integrate) findings from paleoanthropology and related disciplines. More specifically, the emphasis is on paleoenvironmental reconstructions and their importance for formulating adequate hypotheses about our cognitive adaptations. Although the primary focus will be on the so-called *Santa Barbara School*, the aim is broader: to draw lessons and propose guidelines that are equally relevant to other disciplines concerned with the study of human behavioral evolution.

The structure of this chapter is as follows. We begin with an overview of evolutionary psychology, outlining its principal proponents, intellectual precursors, and core tenets. We then examine the methodological framework of the discipline, with particular attention to the so-called top-down and bottom-up approaches, both of which rely in essential ways on paleoanthropological evidence. Next, we examine a specific hypothesis in evolutionary psychology regarding the purportedly evolved human preference for certain habitats, a case that illustrates how claims about our evolutionary past are often advanced without sufficient engagement with the pertinent paleoanthropological and paleoecological record. We show how this hypothesis ought to be revised in light of such evidence and, finally, conclude with broader reflections on how the methodological foundations of evolutionary psychology might be reconfigured to better integrate paleoanthropological insights.

2. Evolutionary psychology: a primer

Evolutionary Psychology (EP) is a discipline that seeks to explain human behavior and cognition in light of the evolutionary processes that shaped the human brain. Its central claim is that the mind is not a general-purpose problem solver, but rather a collection of functionally specialized psychological adaptations (sometimes called *modules*, after Fodor, [1983]), each evolved to solve a recurrent adaptive problem faced by our ancestors. These include mate selection, kin recognition, predator avoidance, social exchange, parenting, and many others.

The intellectual roots of EP trace back to Darwin, ethology, and especially sociobiology, the discipline inaugurated by Edward O. Wilson with the publication of his influential book *Sociobiology: The New Synthesis* (1975)[51]. Wilson applied evolutionary biology to the study of social behavior across animals, including humans, and introduced concepts such as inclusive fitness and kin selection into behavioral research. While sociobiology initially focused mostly on non-human animals, its extension to human behavior was both pioneering and controversial, sparking heated debates about biological determinism, reductionism, and the role of culture in shaping humans' ways of life (see Kitcher, 1985 for a philosophical critique). Despite this controversy, sociobiology laid the groundwork for later approaches by framing behavior as a product of evolutionary processes, and by insisting that natural selection can act not only on morphology but also on behavioral traits, even in humans.

The popularity of sociobiology began to wane when, between the late 1980s and early 1990s, a more cognitively oriented version of this evolutionary approach took shape. A new generation of psychologists and anthropologists, critical of sociobiology's emphasis on behavior as the primary level of analysis (Cosmides & Tooby, 1987), gradually formed into a distinct intellectual movement. This development was driven above all by the work of Leda Cosmides and John Tooby, whose collaborative research—most notably their 1992 contribution to *The Adapted Mind*, co-edited with Jerome Barkow—conferred upon evolutionary psychology its distinctive theoretical identity[52]. Their approach integrated evolutionary biology and adaptationist thinking with information-processing models of the mind, arguing that psychological mechanisms are computational adaptations shaped by natural and sexual selection. According to their view, just as the heart is an organ designed to pump blood, psychological mechanisms are designed to process information in ways that (pre)historically increased fitness. These mechanisms are thought to be predominantly domain-specific, modular, universal, and largely invariant across cultures (Tooby & Cosmides, 1992). The core goal of EP is to uncover the structure and function of these evolved modules and to understand how they manifest in modern behavior, even when modern environments differ markedly from ancestral ones. In doing so, EP aims to contribute to a unified

[51] A detailed examination of the historical precursors of evolutionary psychology lies beyond the scope of this chapter; comprehensive discussions can be found in Brown and Lala (2024) and Griffiths (2008).

[52] Quite early in their career Leda Cosmides and John Tooby moved to Santa Barbara, where they founded the Center for Evolutionary Psychology. Because of this, EP has also been named Santa Barbara School. It is worth bearing in mind that there are other schools of thought that study the evolution of the human mind and could legitimately be described as forms of evolutionary psychology. Historically, however, the label evolutionary psychology has been most closely and consistently associated with the Santa Barbara School. It is to this school that we will refer in this chapter.

scientific understanding of the human mind, informed by both evolutionary theory and cognitive psychology (for recent and in-depth introduction to evolutionary psychology see Buss, 2024).

A key concept in EP is the *Environment of Evolutionary Adaptedness* (EEA), the statistical composite of selection pressures that shaped human cognition. Although in principle each cognitive mechanism may have its own specific EEA, reflecting different selective contexts, many evolutionary psychologists loosely associate our EEA with the African savanna of the Pleistocene (from 2.58 million years ago to 11.700 years ago), where the members of our genus (*Homo*) evolved and lived as hunter-gatherers. What both conceptions have in common is the belief that our cognitive evolution stopped with the invention of agriculture, at the beginning of the Holocene (Tooby & Cosmides, 2016). According to evolutionary psychologists, evolutionary processes are slow, especially when they involve complex organs such as the brain, so it is *a priori* very unlikely that new psychological adaptations have evolved in such a short time (*gradualism*). This means that many of our cognitive mechanisms are not adapted to the contemporary world that arose after the invention of agriculture, and we often experience a mismatch between our psychological capabilities and the demands of the environment. An example of this would be our preference for high-fat or sugary foods: such a preference was probably adaptive in an environment where these resources were scarce and valuable, but maladaptive in today's context where processed and ultra-processed foods are readily available to almost everyone (Buss, 2024).

3. The methods of evolutionary psychology

We briefly presented evolutionary psychology and its core tenets. Now we will introduce EP's two main heuristics for hypotheses generation: the top-down approach and the bottom-up approach. The *top-down* method (also known as functional analysis, see for example Buller, 2005) follows a systematic, theory-driven process that moves from general evolutionary principles, such as inclusive fitness theory, parental investment theory, or sexual selection, to specific hypotheses about the structure and function of cognitive mechanisms.

This approach is articulated in six main steps (Tooby & Cosmides, 1989; Lieberman, Cosmides & Tooby, 2007):

1. *Identification of a recurrent adaptive problem.* The first step involves identifying a problem that would have reliably affected our ancestors' survival or reproduction across evolutionary time. These problems are derived from principles of evolutionary theory and behavioral ecology, and must meet two key criteria: they must have been recurrent in the ancestral environment, and relevant to fitness. Examples include

detecting social cheaters, selecting fertile mates, avoiding predators, or allocating parental resources.

2. *Determine how this adaptive problem manifested in our evolutionary past.* This step is crucial as it grounds hypothesis generation in plausible reconstructions of ancestral conditions. Once an adaptive problem has been defined in abstract terms (such as cheater detection) it is necessary to specify how it would have manifested in our ancestral environment. Reconstructing these contexts provides the necessary constraints for Step 3.
3. *Develop a computational theory that incorporates the information processing problems that need to be solved.* Evolutionary psychologists have to integrate the adaptive problem with what is known about Pleistocene conditions, identify the specific information-processing challenges involved, and use this as a computational model to generate testable hypotheses about the cognitive mechanisms that solve those challenges.
4. *Use the computational theory developed to determine the minimal specifications that a cognitive program must possess to solve the adaptive problem in question, and to construct models of the program's actual structure.* Based on the functional analysis, researchers formulate a hypothesis about a specific cognitive mechanism that may have evolved to solve the problem. This hypothesis includes assumptions about the computational structure of the mechanism: its inputs, domain of application, internal inferential structure, outputs, etc.
5. *Empirical testing of alternative models.* Eliminate erroneous models through experiments (typically through the classic tools of cognitive or social psychology), field observations, and even agent-based modeling, hunter-gatherer studies, archaeology, neuroscience, genetics, and so on.
6. *Assessment of the match between the model and reality*. Compare the predictions provided by the model left with the actual behavior of humans in the modern world. The stronger the match between predicted and observed data, the greater the support for the hypothesized adaptation.

This six-step approach provides a clear logical framework for theory construction in evolutionary psychology, helping to avoid the pitfalls of narrative storytelling, according to Tooby and Cosmides (1989). A canonical example is the cheater-detection module hypothesized by Cosmides in her PhD dissertation (1985; see also Cosmides & Tooby, 1989; 1992). Starting from the adaptive problem of social exchange, and how this problem manifested in our evolutionary past, the theory proceeds through a computational analysis of social exchange in the Pleistocene,

a proposed cognitive adaptation (domain-specific reasoning for social contracts), and predictions on the features of this module (e.g., improved performance on selection tasks involving cheating) that are tested through empirical methods (cognitive experiments and neuroscientific lesion studies, see Cosmides & Tooby, 2016).

The *bottom-up* method, in contrast, reverses this direction of reasoning. Rather than starting from general evolutionary principles and adaptive problems of our past, it begins with the empirical observation of a particular psychological trait, behavior, or preference in the present. The goal is then to infer the possible adaptive function that this trait might have served during our evolutionary past: from the observation of the behavior, researchers attempt to reconstruct the selective context that would have favored such responses. For example, the widespread human fear of snakes, even among individuals who have never had aversive encounters with them, has been interpreted as evidence for a threat-detection mechanisms shaped by the recurrent danger these animals posed in ancestral environments (see Öhman & Mineka, 2001). Since it involves taking into consideration a present-day mechanism and trying to infer the reason for its evolution, this heuristic is often called reverse engineering (Dennett, 1995; Pinker, 1997).

Despite its popularity among evolutionary psychologists, bottom-up reasoning is particularly susceptible to the production of "*just-so stories*"[53], narratives that explain the presence of a trait by appealing to a hypothetical adaptive history, without seeking independent evidence beyond the story's apparent plausibility. Gould and Lewontin (1979) famously argued that adaptationists view organisms as mere assemblages of traits, each independently shaped by natural selection to perform a specific function. Such an approach, they warned, neglects the structural, developmental, and historical constraints that influence evolution. Thus, the risk is mistaking by-products, structural necessities (what they called "spandrels") and other non-selective factors for adaptations. Because bottom-up reasoning starts from traits observed in the present and *assumes* they are adaptations, it risks precisely this error.

On the other hand, the top-down approach would not be subject to this kind of criticism. Indeed, while reverse engineering assumes that a trait is an adaptation and proposes an adaptive explanation for it, functional analysis is predictive, since it is possible to test hypotheses derived from it against the empirical evidence:

> Our hominid ancestors had to be able to solve a large number of complex adaptive problems, and do so with special efficiency. By combining data from paleontology and hunter-gatherer studies with principles drawn from evolutionary biology, one can develop a task analysis that defines the nature of

[53] With reference to Rudyard Kipling's 1902 collection of stories for children, in which the English writer invents amusing and imaginative explanations for how animals acquired certain characteristics.

> the adaptive information-processing problem to be solved. […] Once one understands the nature of the problem, one can then generate very specific, empirically testable hypotheses about the structure of the information processing mechanisms that evolved to solve it. […] One virtue of this approach is that it is *immune to* the usual (but often vacuous) *accusation* of *post hoc storytelling*: The researcher has predicted in advance the properties of the mechanism. (Cosmides et al., 1992, p. 11, emphasis added).

As the quotation above shows, Cosmides, Tooby, and Barkow adopt strikingly strong language, suggesting that functional analysis is entirely *immune* to charges of storytelling. But as we will show shortly, this is not enough to protect functional analysis from accusations of generating just-so stories. Indeed, both top-down and bottom-up approaches often share a deeper methodological limitation: a tendency to operate with minimal integration of paleoanthropological and archaeological evidence. When the Environment of Evolutionary Adaptedness is invoked without careful attention to the actual ecological, technological, and social conditions of ancestral populations, as reconstructed through the historical sciences, evolutionary explanations risk becoming disconnected from plausible scenarios of human evolution and, consequently, being empirically inadequate. But before addressing this specific issue, it is important to consider the several criticisms that have been moved to evolutionary psychology.

4. General methodological and theoretical criticisms

Evolutionary psychology has generated a wealth of hypotheses and empirical work, yet it has also faced sustained criticism from within and beyond evolutionary science (see e.g., Buller, 2006; Richardson, 2007; Pievani, 2014; Nelson, 2017). A persistent concern is the adaptationist bias highlighted by Gould and Lewontin (1979; see also Lloyd, 1999; 2021). A second line of critique targets EP's emphasis on massive modularity and universality: findings in developmental psychology, cognitive neuroscience and even cognitive archaeology indicate substantial plasticity, functional overlap, and developmental reuse (Buller & Hardcastle, 2000; see also d'Errico & Colagè, 2018; Colagè & d'Errico, 2020 for a perspective rooted in archaeology), while cross-cultural research documents wide variation in cognition, norms, and even basic perception (Henrich et al., 2010). Related theoretical developments, such as niche construction theory and gene–culture coevolution, stress that organisms, and humans in particular, modify their environments in ways that reshape selection pressures, with cultural innovations feeding back into biological evolution (Laland et al., 2010). This makes a strictly Pleistocene-fixed view of cognitive architecture less tenable, since evolutionary change continued into the Holocene and involved also genes expressed in the brain that likely influenced our cognitive mechanisms (Mekel-Bobrov et al., 2005; Hawks et al., 2007; Powell, 2012; Hawks, 2014).

Further criticism concerns the reification of the Environment of Evolutionary Adaptedness (EEA) and the limited engagement of evolutionary psychology with paleoanthropological and archaeological evidence. These issues are particularly relevant here, as they are directly connected to the topics discussed in this dissertation, and highlight the need to integrate evidence from these disciplines into EP. The following sections are devoted to exploring how the historical sciences can provide crucial constraints to evolutionary psychological reasoning.

5. The stylization bias

In this section, we address a final critical point that has received relatively little attention in existing literature. As we will argue, the reification of the Environment of Evolutionary Adaptedness, combined with the limited use of paleoanthropological evidence in reverse engineering and functional analyses, has led to the formulation of empirically inadequate hypotheses. This discussion aims to contribute to a more historically grounded approach within evolutionary psychology.

We have already noted that there are two distinct ways of conceptualizing the Environment of Evolutionary Adaptedness (EEA) of the human mind. The first tends to equate it with the Pleistocene African savanna, whereas the second defines it as the set of selection pressures that shaped a given adaptation. Tooby and Cosmides claim that they only ever advocated the latter conception, and only appealed to the former in the early stages of their thinking, to make their ideas more understandable even to those without prior knowledge of evolutionary biology or those who considered all human behavior adaptive in the modern world (Laland & Brown 2011, p. 124). Despite this, a stereotypical image of the EEA—as a Pleistocene savanna inhabited by small bands of hunter-gatherers—appears to have crystallized within certain research traditions of the Santa Barbara School. Combined with the tendency to rely on superficial descriptions of the EEA while largely disregarding recent findings from paleoanthropology and prehistoric archaeology, this has the effect of diverting their heuristic strategies and fostering the development of inadequate hypotheses. We can call this phenomenon "*stylization bias"*, because it involves replacing the most recent findings of historical science with a crystallized and stereotypical view of human evolution. As we will see when examining the environments inhabited by our ancestors, human evolution has been (and continues to be) a complex phenomenon that resists simplistic generalizations. Over time, both our ancestors and the contexts where they lived underwent profound transformations: their habitats, diets, social practices, and much else changed dramatically. For this reason, attempts to reduce such complexity to a single statement—for instance, that "our ancestors lived in the

African savanna"—are inevitably inadequate. In the next section we provide an example of this bias.

6. The savanna hypothesis

Deciding where to move, live and camp is an extremely important adaptive problem for every mobile living organism. Selecting a suitable location—one that offers a reasonable balance between shelter and camping opportunities, access to essential resources (water, food, tool-making materials, and so on), and minimal exposure to predators or other dangers—could mean the difference between life and death. This line of reasoning (step 1 of the top-down approach) led evolutionary psychologists to hypothesize, on the basis that much of our evolution took place in the African savanna (step 2), that modern humans possess a predilection for environments reminiscent of the latter. In other words, a cognitive adaptation expressing an aesthetic preference for savanna-like landscapes where our species evolved. This inclination was then tested through psychological experiments. We are thus faced with a paradigmatic case of functional analysis, where evidence about our evolutionary past serves to provide constraints during the phase of hypothesis generation.

Aptly named the savanna hypothesis[54], this idea was famously proposed by Gordon Howard Orians in a series of articles and book chapters (1980; 1986; Orians, Heerwagen 1992, the latter contribution contained in *The Adapted Mind*, the manifesto of evolutionary psychology). For Orians, the African savanna was an optimal environment for our ancestors: acacia trees provided fruit as well as vantage points of observation and shelter from the sun and predators. At the same time, the fact that the trees were spaced apart, and thus did not form an enclosed environment, allowed the sun's rays to reach the ground, promoting the emergence of grassy expanses on which large herbivores could feed, herbivores on which we began to feed from a certain point of our evolutionary history. The open environment also allowed for a wider field of vision, with obvious benefits in identifying dangers and opportunities. Most importantly, under the "classical" conception of the EEA, the African savanna is portrayed as the environment in which "99 percent of our evolution" took place. According to this view, our lineage evolved and remained in this setting for most of its history until, between roughly 100,000 and 70,000 years ago, *Homo sapiens* left Africa and began to disperse across much of the world (however, a growing body of evidence

[54] This should not be confused with the other "savanna hypothesis", widely debated in paleoanthropology, which argues that the African savanna played a crucial role in shaping the evolution of bipedality in our ancestors (see, for example, Domínguez-Rodrigo 2014 and comments therein). Both versions of the hypothesis, however, have been challenged by recent evidence concerning the paleoecologies of our hominin ancestors.

suggests that waves of *sapiens* migration out of Africa occurred even earlier than this, see, e.g., Hershkovitz et al., 2019). Because our ancestors are thought to have lived in the savanna for the majority of their evolutionary past, it is assumed that humans developed an enduring preference for this type of environment.

Orian's hypothesis was subjected to a series of tests, which at first seemed to confirm its validity. Participants from different regions of the world (the United States, Argentina, and Australia) tended to find acacia trees with moderately dense crowns and trunks that forked close to the ground more attractive (Orians, Heerwagen 1992). Such features were probably useful to our ancestors, as they facilitated climbing and seeking shelter above these trees, and thus would recall to our prehistoric minds the idea of high-quality savannas. The authors noted that depictions of environments reminiscent of tropical savannas are also common in landscape art.

Another study, by Balling and Falk (1982), analyzed the landscape preferences of subjects aged 8 to 70 years, showing images of five biomes (savanna, desert, deciduous forest, tropical forest and coniferous forest). Savanna landscapes were rated as the most desirable (both as places to live and as travel destinations), especially by younger participants, suggesting an innate preference for this type of environment. However, with advancing age, preferences seemed to be influenced by personal experiences, leading to greater parity between the savanna and other forest biomes.

The savanna hypothesis has continued to prove extremely prolific in generating additional tests and case studies (Sommer, 1997; Dutton, 2003; Falk & Balling, 2010). However, the results of these experiments are often ambiguous, and participants do not consistently show a preference for savanna-like environments (e.g., Han, 2007; Hartmann & Apaolaza-Ibáñez, 2010; Moura et al., 2018). Moreover, these findings cannot be fully explained by the idea that people simply learn to prefer the environments in which they were raised, thereby "overriding" an innate preference for the savanna. Indeed, several studies have shown that humans often prefer images of rainforests, even when they have never lived in such environments. (e.g. Moura et al. 2018). Despite this, the savanna hypothesis has had considerable influence in evolutionary psychology and is still presented in the textbooks and works of evolutionary psychologists as a valid explanation for our aesthetic preferences in terms of landscapes (see for example Silverman & Choi, 2015; Buss, 2024).

At this point, it is legitimate to ask why this case study has been chosen for closer examination. After all, isn't it generally accepted that our species evolved primarily on the African savanna? Most depictions of early humans, whether in museum displays, textbooks, or documentaries, tend

to situate them within environments evocative of that biome. The following section examines what paleoanthropological and paleoecological research reveal about this.

7. The paleoecology of our ancestors

Human evolution is sometimes portrayed in popular representations as an "escape" from closed wooded or forested environments, where the common ancestor between humans and chimpanzees probably lived, to the more open ones of African savannas. More open environments are thought to have favored the adoption of an upright posture, which was crucial for freeing the hands to manipulate and make tools. In turn, these tools enabled access to higher-energy resources, thereby supporting the evolution of our brain, a metabolically expensive organ. This process likely created a self-reinforcing, virtuous cycle in which technological, ecological, and cognitive changes mutually amplified one another (Richard Potts dubbed these reconstructions of human evolution "intrinsic explanations", see for example Potts, 1998, for references to several of these reconstructions).

While not entirely mistaken, some aspects of this narrative have been increasingly challenged over the past decades. In particular, the savanna does not appear to have consistently played a decisive role in human evolution. Evidence suggests that some of the earliest hominins lived in closed, wooded, or at least "mosaic" environments, where more closed and more open habitats coexisted, rather than in open savannas. The paleoecology of one the earlier putative hominins, *Orrorin tugenensis* (~6 Mya) points in this direction (Pickford & Senut, 2001). The case is even clearer with *Ardipithecus ramidus*, whose fossils were discovered in Ethiopia and date to about 4.4 million years ago. Paleoenvironmental reconstructions based on its morphology, carbon isotope analyses of dental enamel, associated fauna and flora, and soil isotopic composition all point to a predominantly wooded habitat (see White et al., 2009; White et al., 2015, for reviews, but see Cerling et al., 2011). Moreover, although *A. ramidus* retained many adaptations for arboreal life, it also displayed traits specialized for bipedal locomotion, such as a shortened upper pelvis. This is particularly significant because it shows that around 4 million years ago our ancestors lived in wooded environments and were at least partly arboreal, yet were already exhibiting adaptations for bipedal walking. This evidence weakens the supposed link between bipedalism and the savanna.

Subsequently, this link seems to be consolidating, partly due to a series of climatic changes that promoted the spread of open grasslands across Africa, particularly in East Africa, a region

central to several major developments in human evolution (Cerling, 1992; Cerling et al., 1997)[55]. Accordingly, members of the genus *Australopithecus* appear to be better adapted to arid environments than *Ardipithecus*. But even during this period, evidence indicates that hominins inhabited mosaic environments combining features of both forests and savannas. Moreover, these long-term aridification trends were punctuated by alternating wet and dry phases, which caused significant fluctuations in vegetation patterns (Levin, 2015).

However, this seemingly privileged link with the African savanna is disrupted by *Homo ergaster/Homo erectus*, the first hominin to leave Africa and colonize parts of Asia (Rightmire, 2001). The reasons behind these migrations are still not entirely clear, but during its dispersals *H. ergaster* encountered a wide range of environments beyond the savanna, from tropical forests (Roberts et al., 2016) to highly arid deserts (Mercader et al., 2025). This demonstrates that *H. erectus* already possessed considerable ecological flexibility that, until recently, was thought to characterize only later members of our genus, such as the Neanderthals and *Homo sapiens*. Subsequently, *Homo heidelbergensis*, which emerged during the Middle Pleistocene (Profico et al., 2016), is thought to have been the main actor in a second major dispersal out of Africa, spreading into Europe and Asia. From these populations, Neanderthals in Eurasia and likely Denisovans in eastern Asia later evolved and developed specific adaptations to cold climates (Holliday, 1997; Huerta-Sánchez et al., 2014).

But it was primarily our species that inhabited and exploited a wide range of biomes—from open savannas and temperate woodlands to coastal areas and high-altitude environments—including tropical forests, which were also widespread in Africa, and remain poorly studied (Scerri et al., 2022)[56]. During the Pleistocene and the Holocene, African forests went through periods of expansion and contraction driven by climatic fluctuations. In many regions, this resulted in mosaic landscapes where forests alternated with savannas and grasslands. These mixed environments offered unique opportunities for hominins, allowing them to exploit the resources of both ecosystems. Far from being hostile places to avoid, as sometimes still assumed, these environments instead represented resource-rich contexts (Roberts & Petraglia 2015). For example, the site of Panga ya Saidi (Kenya) shows that humans produced symbolic tools, practiced burials and exploited mixed forest environments 78 ka (Shipton et al. 2018, Martinón-Torres et al., 2021). With the outflow of *sapiens* from Africa, we witness within not even 100,000 years the

[55] The climatic changes were caused by several factors, including the tectonic uplift of East Africa, which led to a reorganization of atmospheric circulation (Sepulchre et al., 2006).

[56] There are several reasons for this. First, fossils are less likely to be preserved in tropical environments, where heat, humidity, and acidic soils accelerate decomposition. Second, Western researchers and colonizers have long regarded tropical forests as hostile, untouched wilderness, which has contributed to a lasting bias against studying them (Patrick Roberts, personal communication).

colonization of almost every environment on planet Earth, confirming our incredible adaptive abilities (Westaway et al. 2017 reports one of the oldest non-African sites where sapiens are attested in a tropical forest, while we refer to the article by Scerri et al. 2022 for an overview of the most recent findings on the occupation of tropical African and non-tropical forests during prehistoric times).

Finally, the great climatic variability of the Pleistocene should also be considered. Indeed, throughout the Pleistocene there was an alternation of cold and dry phases and warm and wet phases, which caused a series of profound environmental changes (Lisiecki & Raimo, 2005; Herbert, 2023). Climatic fluctuations probably favored the selection of some behavioral flexibility, rather than adaptation to a specific biome. Adaptability to different environmental conditions is also highlighted by the various waves of migrations that our ancestors underwent, the first of which dates to *ergaster/erectus*. Building on this, several theoretical models try to explain the relationship between the evolution of our ancestors and the environment. In recent years, the "variability selection hypothesis", proposed by Richard Potts, has had particular resonance (e.g., 1998; 2012; see Vrba, 2015, for a complementary account of the relationships between human evolution and the unstable climatic and ecological conditions). According to Potts, fundamental changes in human evolution occurred concurrently with these phases of increased instability and phase reversal, and have increased the ecological flexibility and adaptability of our ancestors to a range of different environments (see also Parravicini & Pievani, 2016, for an overview of the role of ecological patterns in human evolution).

In summary, human evolution did not unfold exclusively in open environments such as the savanna; rather, our ancestors were capable of dispersing into and colonizing a wide range of habitats. Therefore, while the savanna undoubtedly played an important role in our evolutionary history, there is no compelling reason to regard it as our preferred biome or as the one in which we felt safest. At the present state of knowledge, it does not appear that most human adaptations have evolved to respond to the demands of a particular environment, so much as to respond flexibly and adaptively to a very variable ecological context (to further explore the wide range of environments in which our ancestors lived, see Foister et al. 2023; Zeller et al. 2023).

In light of the evidence presented, the savanna hypothesis, as intended by evolutionary psychologists, is clearly inadequate. The heuristic process (the functional analysis) that generated it is based on a stereotypical view of human evolution that equates our evolutionary past with life in African savannas, and remains largely informed by paleoecological data. Recall that functional analysis starts by identifying an adaptive problem—in this case, determining the ideal habitat for human survival (step 1)—and then reconstructing how that problem manifested in our evolutionary

past (step 2). It is precisely at this stage that the main difficulties arise when it comes to integrating paleoanthropological and paleoenvironmental evidence. As we have seen, most evolutionary psychologists argue that the African savanna provided the primary stage for human evolution. From this assumption, they infer that we should have developed an enduring preference for such environments, a preference that should still be observable in humans today. While this reconstruction may have seemed plausible when the hypothesis was first proposed, it is no longer tenable in light of paleoanthropological evidence, which now points to a far more complex evolutionary scenario. Unfortunately, the tendency of evolutionary psychologists to underutilize paleoanthropological evidence and to rely on overly simplified reconstructions of our past, what we deemed the *stylization bias*, has allowed the savanna hypothesis to persist, albeit with mixed results. The inconsistent results of landscape-preference experiments can be better understood by recognizing that human evolution occurred in diverse habitats, shaped by adaptability and ecological versatility rather than specialization in a single biome.

But how should we revise the functional analysis underlying the savanna hypothesis in light of these remarks? Step 1 is not controversial in itself; choosing the right place to live is crucial for survival. The problem arises with step 2, which should draw on all the relevant evidence provided by paleoanthropology and, in this case, paleoenvironmental reconstructions. Rather than relying on a stereotyped account of human evolution, we should acknowledge its complexity and the environmental diversity in which our ancestors lived. If we were to preserve the same inferential structure as the original model, we might instead expect humans to show a preference for mosaic environments, a conclusion also reached by Rathmann and colleagues (2022), or, more generally, for natural settings over heavily anthropized ones such as large cities. Alternatively, this preference could reflect a broader adaptive strategy for identifying suitable habitats without privileging any specific environment[57].

8. How to revise functional analysis

In section 4, when presenting the different heuristics of hypothesis generation, we showed that the founders of evolutionary psychology regard the top-down method as immune to accusations of producing just-so stories. This supposed immunity stems from the method's predictive nature:

[57] One might argue that the distinction between savanna and forest is not always clear-cut, since the main defining feature of the savanna is simply the presence of a grassy substrate. On this view, savannas could encompass a wide variety of environments, including mosaic landscapes with substantial tree cover (see the paragraph "What is a savanna?" in Dominguez-Rodrigo, 2014). This is a fair point; however, it is not the conception of savanna adopted by proponents of the savanna hypothesis in evolutionary aesthetics. In their studies, participants are typically presented with images of environments characterized by sparse tree cover. This represents yet another instance in which stylization bias has shaped and, ultimately, misled research.

it begins with the formulation of a hypothesis about a cognitive adaptation, which is then tested against empirical evidence, thereby avoiding the construction of *ad hoc* explanations. In the last three sections, we have shown that this immunity is far from guaranteed: if evidence from our past is not properly integrated into the process of hypothesis generation, there is a risk of producing empirically inadequate hypotheses. One might object that, in functional analysis, hypotheses are in any case tested against empirical evidence, and therefore can be refined or corrected through this process. This objection would be valid if evolutionary psychologists were willing to revise their hypotheses in light of empirical testing[58]. But this is not what happened with the savanna hypothesis. Despite mixed results, the hypothesis continues to be tested and appears in introductory texts on evolutionary psychology.

Because of the stylization bias, EP hypothesis, compared to those of paleoanthropology, have typically a low *spatial*, *temporal* and *evolutionary* "resolution": they often treat the entire evolutionary history of humankind as a single, homogeneous block, overlooking the numerous transformations our lineage has undergone across geography, time and phylogeny—from the succession of different species, to shifting environmental and ecological conditions, to the growing diversity and complexity of human lifeways. Paleoanthropology and archaeology, on the other hand, operate at a much finer resolution. Although EP must to some extent abstract from certain details to keep the generation process functional (it would not be feasible to consider *all* the evidence about our past), the level of abstraction caused by stylization bias leads to a hypothesis generation process that is inadequate.

But how should evolutionary psychologists revise functional analysis? We have seen that Tooby and Cosmides (1989) propose a process with six steps. In this partition of functional analysis, historical evidence is relevant to the second step: *identification of a recurrent adaptive problem* and *characterize how the adaptive problem manifested itself in our evolutionary past.* For these steps to be successfully completed, it is necessary to adequately reconstruct the EEA of an adaptation, that is, the set of selective pressures that shaped it. To do so, we should ideally do two things:

1. *Outline the period of interest*. As already mentioned, not all adaptations will have the same EEA. To identify the relevant period, an initial review of paleoanthropological

[58] Another related issue raised by critics is the weakness of the results of tests conducted by evolutionary psychologists. See Buller (2005) for an in-depth examination of the flaws in some of the most famous experimental paradigms, such as cheater detection and mate selection. Nelson (2017) and Dennett (1995), on the other hand, point out that evolutionary psychologists fail to consider and test alternative hypotheses to adaptationist ones (see also Gray et al., 2003, for criticisms related to the research on the waist-to-hip ratio).

literature is already necessary[59]. For example, meat consumption plays a fundamental role in hypotheses about the evolution of disgust (Kelly, 2011): the evolution of this primordial emotion would be linked to proximity to carcasses, as scavengers and/or predators. But when did the systematic consumption of meat begin? In other words, what is the EEA of disgust? Evolutionary psychology alone cannot answer this question; it needs paleoanthropological evidence.

2. *Select all available relevant evidence*. This step clearly depends on the adaptive problem we are investigating. If, for example, we are investigating the assumed differences in partner choice strategies between men and women, any data about the possible social structure of our ancestors and their evolution over time is pertinent (more on this later). Number of individuals per group, sexual dimorphism, ratio between males and females, the presence of different roles, group mobility and intra-group migration. But also patterns of development and other data that can give us clues as to how the burden of childcare was distributed, and many more[60].

Of course, point 2 is not straightforward. Our knowledge of the past is extremely fragmentary, but it is rapidly increasing and being revised, which makes it difficult to stay up to date when conducting research in another discipline. At the same time, understanding what is relevant for reconstructing the EEA of a particular adaptation is not an easy task. In this regard, greater interdisciplinarity and collaboration with paleoanthropologists and archaeologists could be beneficial in this regard. Evolutionary psychologists often claim that the overarching aim of their discipline is the unification of all the human evolutionary sciences (e.g., Cosmides & Tooby, 2016). Regardless of whether this undertaking is ultimately feasible, incorporating evidence from these disciplines appears to be one of the minimum requirements for pursuing this goal.

Historical evidence plays a similar role in reverse engineering. In this case, we start with the observation of a behavioral or psychological pattern in modern humans, but knowledge of the past is still required to work out the adaptive function of that particular trait. What is the function of jealousy (if any)? We cannot know unless we reason in terms of the specifics of the context in which it would have evolved. Moreover, paleoanthropological evidence is relevant for another reason. We mentioned that, according to Gould and Lewontin (1979), adaptationist bias improperly leads researchers to divide organisms into independent traits. This issue also arises in evolutionary

[59] If the EEA of a particular adaptation for example calls into question primate evolution more generally (think of the attachment relationship between mother and child) then evidence from primate paleontology and primatology also become relevant.

[60] Historical evidence is relevant for the following steps as well, as they require details of what informational resources were available to our ancestors, to infer what specifics of the cognitive mechanism evolved.

psychology, where the choice of a particular character as a trait with a separate evolutionary history is rarely justified. Considering the archaeological and fossil record can help evolutionary psychologists reconstruct the history of certain characters and determine which are genuine individual traits and which are not. For example, Hauser et al. (2002) proposed that the language faculty should not be regarded as a single character but as a mosaic of different traits with distinct evolutionary histories—a view consistent with the fossil and archaeological record (see Parravicini & Pievani, 2018, 2019). For this reason, an evolutionary psychologist applying reverse engineering to the language faculty may actually be analyzing a composite of different characters, and thus working at the wrong level of analysis.

One final consideration is in order: fully integrating the evidence about our past risks being deeply disruptive to the methods of evolutionary psychology. The increasing complexity of reconstructions of our evolution seems to indicate a certain instability in the selective pressures to which we have been subjected, especially with regard to certain aspects of group life and adaptation to the environment. Identifying a cheater, collaborating with one's companions, finding the best place to set up camp, recognizing environmental dangers, etc. are problems that have changed continuously throughout our evolution. For this reason, it is difficult to imagine that natural selection has favored a set of fixed modules rather than a certain cognitive flexibility (a perspective shared by several critics, see for example Sterelny & Griffiths, 1999). The inclusion of paleoanthropological evidence, therefore, risks being incompatible with the principles of the Santa Barbara School. Unless evolutionary psychologists are willing to admit that our cognitive mechanisms are more flexible and adaptable than they usually acknowledge.

9. Conclusions

We have seen that, despite its relevance, evidence from paleoanthropology and archaeology is not always integrated into disciplines studying human cognitive and behavioral evolution. Focusing on evolutionary psychology and the savanna hypothesis, we have shown that its formulation overlooks paleoenvironmental evidence about the contexts of human evolution. Consequently, evolutionary psychologists may be testing empirically inadequate hypotheses[61]. Although there appear to be tentative signs of renewal within EP (e.g., Barrett, 2015; see also Brown & Lala, 2024), this remains an issue that evolutionary psychologists must address. Indeed, some evolutionary psychologists seem to have recently taken a self-critical stance, admitting that

[61] To be clear, we are not claiming that every hypothesis in evolutionary psychology suffers from this problem; but rather that *if* historical evidence is not taken into account during hypothesis generation, *then* we risk formulating inadequate hypotheses.

the discipline should take archaeological and paleoanthropological evidence more seriously (e.g. Fessler et al. 2016; Abramiuk, 2021). And ultimately, evolutionary psychologists cannot get away from doing this if they really want to achieve that unification of multiple fields of knowledge that they so yearn for.

The same concerns also apply to other related disciplines. As already mentioned, Meneganzin and Currie (2025) recently emphasized that human behavioral ecology needs to integrate evidence from paleoanthropology in order for its hypotheses to be historically sound. Others have expressed the importance that models of cultural evolutionary theory be grounded in and tested against the archaeological record (e.g. Garvey, 2018; Tostevin, 2019). In any case, all disciplines concerned with the evolution of the human mind must engage with evidence from paleoanthropology and archaeology. These fields have long drawn on insights and methods from other sciences, yet they also have much to contribute in return.

CONCLUSIONS

This dissertation has examined evidential reasoning in paleoanthropology and Paleolithic archaeology, with the aim of clarifying how fragmentary and degraded traces of the past are transformed into evidence for claims about hominin behavior, cognition, and evolution. Rather than focusing exclusively on the availability of data, the thesis has emphasized the inferential practices through which data are mobilized, evaluated, and communicated within the human evolutionary sciences.

A first contribution of this work is conceptual. By analyzing the debate surrounding the alleged burials of *Homo naledi*, the thesis has articulated two distinct conceptions of evidential standards: a data-checklist view, according to which standards primarily specify what kinds of data must be collected, and a scaffolding view, according to which standards concern the soundness of the inferential structure linking data to claims. The analysis has shown that disagreement in the *Homo naledi* case does not hinge on which standards are relevant—these are largely shared—but on how those standards are interpreted and operationalized. This distinction helps explain why appeals to evidential standards can coexist with deep and persistent disagreement, and it provides a general framework for diagnosing similar disputes in other areas of the historical sciences.

The analysis is also extended to standards of communication, emphasizing proportionality and inferential disclosure as conditions for responsible scientific dissemination in "sensational science".

A second contribution concerns the epistemic role of variability in Paleolithic mortuary practices. By introducing the variability-of-expression challenge, the thesis has argued that variability is a structural feature of the archaeological record that places substantive constraints on evidential reasoning. This challenge raises issues both for our definition of what constitutes a burial and for our practices aimed at identifying burials in the archaeological record. The analysis showed that variation poses a problem of detection: how to define what counts as a burial, what does not, and on what grounds. With respect to identifying burials in the archaeological record, rigid diagnostic lists of criteria (such as the one endorsed by the *naledi* team) and null-hypothesis reasoning perform poorly under such variability, with the risk of entrenching *sapiens*-centric standards, or, conversely, of introducing overly idiosyncratic ones. By contrast, criteria interpreted inferentially, together with comparative and explanatory approaches, can properly address the variability-of-expression challenge, granted that they are calibrated with rigor. In summary, no single strategy can fully neutralize the effects of variability.

The third contribution is a critical assessment of the use of null hypotheses in cognitive archaeology. Two recent case studies and two different usages of null hypotheses in cognitive archaeology are presented: pseudo-nulls and heuristic nulls. While the first strategy has already been discussed in philosophy of science, the second is acknowledged for the first time in this chapter. It is shown that importing statistical null-hypothesis reasoning into non-experimental, historical contexts generates heuristic nulls that suffer from structural problems of testability. By contrast, inference to the best explanation better captures how archaeological claims are actually supported, namely through the convergence of multiple, independently fragile lines of evidence. This analysis clarifies why some methodological tools that are indispensable in experimental sciences become problematic when transposed to the study of the deep past.

A further contribution of this dissertation concerns the integration between paleoanthropology and disciplines that aim to explain the evolution of the human mind. Focusing on evolutionary psychology, the analysis shows that hypotheses about evolved cognitive adaptations are often generated and assessed with insufficient engagement with paleoanthropological and archaeological evidence. This results in empirically underconstrained narratives, as illustrated by the savanna hypothesis. By introducing the notion of stylization bias, the thesis clarifies how simplified reconstructions of the Environment of Evolutionary Adaptedness lower evidential standards and facilitate just-so explanations. The analysis argues that evolutionary psychology's top-down and bottom-up heuristics implicitly rely on historical assumptions that remain largely unexamined. More generally, the chapter shows that historically grounded evidence should function as a constraint on hypothesis generation, rather than as optional background information, thereby extending the thesis' account of evidential reasoning beyond paleoanthropology proper.

Conceptually, the dissertation (i) clarifies the nature of evidential standards in historical sciences and their different interpretations; (ii) extends the discussion to norms of scientific communication; (iii) articulates the variability-of-expression challenge for the definition and identification of burials in funerary archaeology; (iv) distinguishes pseudo-nulls, heuristic nulls, and genuine null usages, offering inference to the best explanation as a corrective for cognitive archaeology; and (v) articulates the stylization bias in evolutionary psychology, and illustrate how existing heuristics might be revised to integrate paleoanthropological and archaeological evidence.

Taken together, these analyses illuminate different facets of evidential reasoning in paleoanthropology and archaeology. In line with previous reflections in the philosophy of the historical sciences, this dissertation has shown that in fields such as paleoanthropology, epistemic flexibility and contextual adaptability are key to addressing the challenges of a fragile evidential base. This flexibility is particularly important when dealing with signs of complex behavior in

different species, because this phenomenon raises a series of questions for our standards. Should we calibrate them to *sapiens*, at the risk of setting too high a bar for other species? Or should we adopt broader criteria that apply to a wider range of species but risk being overly permissive? Or even adapt our criteria to different contexts, at the cost of turning them into overly idiosyncratic standards? Finding answers to these questions is by no means easy, and, as we have seen, many debates in paleoanthropology continue to revolve around these issues. For this reason, it may be especially fruitful to pursue further philosophical reflection on evidential standards: on their nature, development, and possible reinterpretation.

Another theme that emerged is integration: historical scientists are often praised for their ability to opportunistically borrow resources from other disciplines to expand their explanatory reach. As is often pointed out, this integration of resources is usually a long and complicated process. But, over the course of the chapters, a new aspect also seems to have emerged: sometimes, as in the case of null hypotheses, integration is bound to fail. Borrowing from other disciplines does not always pay off. It is recommended that future research focus on these issues. In philosophy of science, the emphasis is sometimes on "success stories", but it can also be useful to analyze cases where the integration of technologies, methodologies, and knowledge from other disciplines does not work. It is important to note that such situations can also provide insights into the requirements for the fruitful borrowing of resources. Furthermore, in the last chapter, we saw that it may also be useful to consider the reverse process: paleoanthropology, archaeology, and related disciplines can "return the favor" by lending some of their insights to other fields. This is also a phenomenon that would deserve further investigation.

And what about our initial concern, pessimism? As we have seen, pessimism in paleoanthropology assumes multiple forms. A comprehensive response to all these forms would require a more systematic and dedicated inquiry. Nonetheless, some preliminary considerations can be drawn. The case of the debate surrounding *Homo naledi* is very instructive. On the one hand, it illustrates that some of the concerns mentioned in the introduction are indeed well-founded. Paleoanthropology is a "superstar science", as evidenced by the widespread resonance of claims about the alleged burials and engravings of *H. naledi* and by the haste to publish results that were still partial. Furthermore, as reported, numerous media headlines repeated the familiar cliché of a "new discovery that rewrites human evolution"[62].

[62] Although it must be acknowledged that, even without rewriting the entire reconstruction of human evolution, the adoption of complex behaviors by *naledi* would still have a significant impact on our theories regarding the relationship between cranial capacity and symbolic behaviors and on the phylogeny of the latter.

On the other hand, there are also reasons for optimism. The response from the scientific community was prompt and cohesive, as discussed in the first chapter. Paleoanthropologists and archaeologists were quick to criticize and to enforce a realignment with the evidential standards shared by the wider community. This was, of course, not sufficient to prevent the publication of the second round of preprints as versions of record. However, this situation opens up possible future directions for the work developed in this thesis. The naledi case remains a matter of ongoing debate, with both the discovery team and its critics continuing to engage in discussion.

Another aspect that deserves evaluation is the influence of the new publication model adopted by *eLife* on the evolution of the *naledi* debate. To what extent, if at all, has this model encouraged the publication of provisional results? How might public review reshape internal discussions within the discipline? And above all, how has it affected the shifts in argumentation between the first and second preprints? These are questions that must be addressed if, as seems likely, the new publication model adopted by *eLife* is here to stay (and may even be embraced by other journals). It is not far-fetched to think that such innovations in scientific publishing could profoundly influence how research is conducted in these disciplines.

Finally, we would like to stress that an understanding of paleoanthropology and archaeology (and of human evolution more in general) has become increasingly relevant across other branches of philosophy. It is now difficult to engage in discussions of ethics, political philosophy, or philosophical anthropology without taking into account fossil, archaeological, and genetic evidence. Questions about what it means to be human can no longer rest solely on *a priori* philosophical reflection; they must also consider empirical findings concerning the evolution of tool use, hybridization with other species of our genus, shifts in developmental patterns throughout our phylogeny, our reciprocal relationship with ecological and climatic changes, and many others. Hopefully, the epistemological analysis of the human evolutionary sciences will become increasingly relevant within general philosophy.

In closing, the path “from bones to beliefs” is neither a leap of faith nor a mere accumulation of facts. It is the disciplined practice of constructing, revising, and communicating inferential bridges from traces to claims, guided by standards that are constantly negotiated. Seen in this light, the historical sciences are not epistemically second-class; they are laboratories of methodological creativity, where rigor lies in building scaffolds strong enough to bear weight and moldable enough to be rebuilt when better designs emerge.

BIBLIOGRAPHY

Abramiuk, M. A. (2021). Evolutionary Psychology and Archaeology. *The SAGE Handbook of Evolutionary Psychology. Integration of Evolutionary Psychology with other Disciplines*, 257-282.

Alexeeva, T. and Bader, N. (eds) (1998) *Upper Palaeolithic Site Sungir: Graves and Environment*. Moscow: Scientific World.

Allen, C. (2006). Transitive inference in animals: Reasoning or conditioned associations? In S. Hurley & M. Nudds (Eds.), Rational animals? (pp. 175–185). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528272.003.0007

Anderson, J. R. (2016). Comparative thanatology. Current Biology, 26(13), R553-R556.

Anderson, J. R. (2017). Comparative evolutionary thanatology of grief, with special reference to nonhuman primates. *Japanese Review of Cultural Anthropology*, 18(1), 173-189.

Anderson, J. R. (2018). Chimpanzees and death. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1754), 20170257.

Anderson, J. R., Biro, D., & Pettitt, P. (2018). Evolutionary thanatology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1754), 20170262.

Andrews, K. (2020). How to study animal minds. Cambridge University Press.

Andrews, K. (2024). "All animals are conscious": Shifting the null hypothesis in consciousness science. *Mind & Language*, *39*(3), 415-433.

Andrews, K., & Huss, B. (2014). Anthropomorphism, anthropectomy, and the null hypothesis. Biology & Philosophy, 29, 711-729. doi: https://doi.org/10.1007/s10539-014-9442-2Bausman, W., & Halina, M. (2018). Not null enough: pseudo-null hypotheses in community ecology and comparative psychology. *Biology & Philosophy*, 33(3), 30.

Aujoulat, N., Geneste, J.-M., Archambeau, C., Delluc, M., Duday, H. and Gambier, D. (2001) La Grotte ornée de Cussac (Dordogne), observations liminaires. *Paléo* 13: 9–18.

Aujoulat, N., Geneste, J.-M., Archambeau, C., Delluc, M., Duday, H. and Henry-Gambier, D. (2002) La Grotte ornée de Cussac – Le Buisson-de-Cadouin (Dordogne): premières observations. *Bulletin de la Société Préhistorique Française* 99(1): 129–37.

Balling, J. D., & Falk, J. H. (1982). Development of visual preference for natural environments. *Environment and behavior*, *14*(1), 5-28.

Balzeau, A., Turq, A., Talamo, S., Daujeard, C., Guérin, G., Welker, F., ... & Gómez-Olivencia, A. (2020). Pluridisciplinary evidence for burial for the La Ferrassie 8 Neandertal child. *Scientific reports*, *10*(1), 21230.

Barham, L., Duller, G. A., Candy, I., Scott, C., Cartwright, C. R., Peterson, J. R., ... & Nkombwe, P. (2023). Evidence for the earliest structural use of wood at least 476,000 years ago. *Nature*, *622*(7981), 107-111.

Barley, N. (1997). Grave matters: a lively history of death around the world.

Barrett, H.C., (2015). *The Shape of Thought: How mental adaptations evolve*. Oxford University Press, Oxford.

Bedetti, M. (2025). *La mente preistorica. Indagine epistemologica sulla psicologia evoluzionistica*. Mimesis Editore: Milano.

Bedetti, M., & Allen, C. (2025). Advancing paleoanthropology beyond default nulls. *Behavioral and Brain Sciences*, *48*, e3. Commentary on Stibbard-Hawkes's target article "Reconsidering the link between past material culture and cognition in light of contemporary hunter–gatherer material use." Behavioral and Brain Sciences 48 (2025): e1.

Been, E., Hovers, E., Ekshtain, R., Malinski-Buller, A., Agha, N., Barash, A., ... & Barzilai, O. (2017). The first Neanderthal remains from an open-air Middle Palaeolithic site in the Levant. *Scientific Reports*, *7*(1), 2958.

Been, E., & Barzilai, O. (2024). Neandertal burial practices in Western Asia: How different are they from those of the early Homo sapiens?. *L'Anthropologie*, *128*(3), 103281.

Belfer-Cohen, A., Bar-Yosef, O., 2000. Early sedentism in the Near East: a bumpy ride to village life. In: Kuijt, I. (Ed.), Life in Neolithic farming communities: social organization, identity, and differentiation. Kluwer Academic Publishers, New York, http://dx.doi.org/10.1007/0-306-47166-3_2, pp. 19–38.

Belfer-Cohen, A., & Hovers, E. (1992). In the eye of the beholder: Mousterian and Natufian burials in the Levant. *Current Anthropology*, *33*(4), 463-471.

Berger LR, Hawks J. *Cave of bones: A true story of discovery, adventure, and human origins*. Johannesburg: Penguin Random House; 2023.

Berger, L. R., Hawks, J., de Ruiter, D. J., Churchill, S. E., Schmid, P., Delezene, L. K., ... & Zipfel, B. (2015). Homo naledi, a new species of the genus Homo from the Dinaledi Chamber, South Africa. *elife*, *4*, e09560.

Berger Lee R, Makhubela Tebogo, Molopyane Keneiloe, Krüger Ashley, Randolph-Quinney Patrick, Elliott Marina, Peixotto Becca, Fuentes Agustín, Tafforeau Paul, Beyrand Vincent, Dollman Kathleen, Jinnah Zubair, Brewer Gillham Angharad, Broad Kenneth, Brophy Juliet, Chinamatira Gideon, Dirks Paul H. M., Feuerriegel Elen, Gurtov Alia, Hlophe Nompumelelo, Hunter Lindsay, Hunter Rick, Jakata Kudakwashe, Jaskolski Corey, Morris Hannah, Pryor Ellie, Ramaphela Maropeng, Roberts Eric, Smilg Jacqueline S., Tsikoane Mathabela, Tucker Steven, van Rooyen Dirk, Warren Kerryn, Wren Colin D., Kissel Marc, Spikins Penny, Hawks John (2023) Evidence for deliberate burial of the dead by Homo naledi eLife 12:RP89106. https://doi.org/10.7554/eLife.89106.1

Berger Lee R., Hawks John, Fuentes Agustin, van Rooyen Dirk, Tsikoane Mathabela, Ramalepa Maropeng, Nkwe Samuel, Molopyane Keneiloe (2023) 241,000 to 335,000 Years Old Rock Engravings Made by Homo naledi in the Rising Star Cave system, South Africa eLife 12:RP89102. https://doi.org/10.7554/eLife.89102.1

Lee R Berger, John Hawks, Agustín Fuentes, Dirk Van Rooyen, Mathabela Tsikoane, Maropeng Mpete, Samuel Nkwe, Keneiloe Molopyane (2025) An initial report of circa 241,000- to 335,000-year-old rock engravings and their relation to Homo naledi in the Rising Star cave system, South Africa eLife 12:RP89102.

Bermúdez de Castro, J. M., Martinón-Torres, M., Carbonell, E., Sarmiènto, S., Rosas, A. *et al*. (2004) The Atapuerca sites and their contribution to the knowledge of human evolution in Europe. *Evolutionary Anthropology* 13: 25–41.

Binford, L. R. (ed.) (1977), For Theory Building in Archaeology: Essays on Faunal Remains, Aquatic Resources, Spatial Analysis and Systemic Modeling, New York: Academic Press.

Bokulich, A., & Parker, W. (2021). Data models, representation and adequacy-for-purpose. *European Journal for Philosophy of Science*, 11, 1-26.

Bower Bruce 2022 Bower, Bruce. 2022. “Homo Naledi May Have Lit Fires in Caves at Least 236,000 Years Ago.” December 2, 2022. https://www.sciencenews.org/article/homo-naledi-fire-hominid-cave-human-evolution.

Bowler, J. M., Jones, R., Allen, H., & Thorne, A. G. (1970). Pleistocene human remains from Australia: a living site and human cremation from Lake Mungo, western New South Wales. *World archaeology*, *2*(1), 39-60.

Bramble, D. M., & Lieberman, D. E. (2004). Endurance running and the evolution of Homo. *nature*, *432*(7015), 345-352.

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis*, *73*(3), 295-311.

Brigandt, I. (2012). The dynamics of scientific concepts: The relevance of epistemic aims and values. *Scientific concepts and investigative practice*, *3*, 75-103.

Brophy, J. K., Elliott, M. C., De Ruiter, D. J., Bolter, D. R., Churchill, S. E., Walker, C. S., Hawks, J., & Berger, L. R. (2021). Immature Hominin Craniodental Remains From a New Locality in the Rising Star Cave System, South Africa. PaleoAnthropology. https://doi.org/10.48738/2021.iss1.64

Brown, G. R., & Lala, K. N. (2024). *Sense and nonsense: Evolutionary perspectives on human behaviour*. Oxford University Press.

Brumm, A., Aziz, F., Van den Bergh, G. D., Morwood, M. J., Moore, M. W., Kurniawan, I., ... & Fullagar, R. (2006). Early stone technology on Flores and its implications for Homo floresiensis. Nature, 441(7093), 624-628.

Bruner, E. & Moggi Cecchi, J. (2022). Encefalizzazione. In Sineo, L. & Moggi Cecchi, J. (Eds.) *Manuale di Antropologia. Evoluzione e biodiversità umana*. Utet: Torino.

Buller, D.J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. MIT press, Cambridge.

Buller, D.J., & Hardcastle, V. (2000). Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind*, *1*(3), 307-325.

Bunn, H. T., & Pickering, T. R. (2010). Bovid mortality profiles in paleoecological context falsify hypotheses of endurance running–hunting and passive scavenging by early Pleistocene hominins. *Quaternary Research*, *74*(3), 395-404.

Buskell, A., & Currie, A. (2021). Uniqueness in the life sciences: how did the elephant get its trunk?. *Biology & Philosophy*, *36*, 1-24.

Buss, D. (2024). Evolutionary Psychology: The New Science of the Mind, 7th edition. Routledge, London.

Campanaro, D. M. (2021). Inference to the best explanation (IBE) and archaeology: Old tool, new model. *European Journal of Archaeology*, *24*(3), 412-432.

Cardini, L. 1980. La necropoli mesolitica delle Arene Candide (Liguria). *Memorie dell'Istituto Italiano di Paleontologia Umana*, n.s. 3:9–31.

Carlson, K. J., Grine, F. E., & Pearson, O. M. (2007). Robusticity and sexual dimorphism in the postcranium of modern hunter-gatherers from Australia. *American Journal of Physical Anthropology*, 134, 9–23.

Cartmill, M. (2002). Paleoanthropology: science or mythological charter?. *Journal of anthropological research*, *58*(2), 183-201.

Cartwright, N. (2007). Are RCTs the gold standard?. BioSocieties, 2(1), 11-20.

Chapman, R., & Wylie, A. (2016). *Evidential reasoning in archaeology*. Bloomsbury Publishing.

Cerling, T. E. (1992). Development of grasslands and savannas in East Africa during the Neogene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *97*(3), 241-247.

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., & Ehleringer, J. R. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature*, *389*(6647), 153-158.

Cerling, T. E., Wynn, J. G., Andanje, S. A., Bird, M. I., Korir, D. K., Levin, N. E., ... & Remien, C. H. (2011). Woody cover and hominin environments in the past 6 million years. *nature*, *476*(7358), 51-56.

Clark, G. A. (2000). On the Questionable Practice of Invoking the Metaphysic. *American Anthropologist*, *102*(4), 851–853. http://www.jstor.org/stable/684211.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. Philosophy of science, 69(3), 474-496.

Cleland, C. E. (2011). Prediction and explanation in historical natural science. The British Journal for the Philosophy of Science.

Colagè, I., & d'Errico, F. (2020). Culture: The driving force of human cognition. *Topics in cognitive science*, *12*(2), 654-672.

Coolidge, F. L., Haidle, M. N., Lombard, M., & Wynn, T. (2016). Bridging theory and bow hunting: human cognitive evolution and archaeology. *Antiquity*, *90*(349), 219-228.

Corbey, R., Jagich, A., Vaesen, K., & Collard, M. (2016). The acheulean handaxe: More like a bird's song than a beatles' tune?. Evolutionary Anthropology: Issues, News, and Reviews, 25(1), 6-19.

Corneille, O., Havemann, J., Henderson, E. L., IJzerman, H., Hussey, I., Orban de Xivry, J. J., ... & Lotter, L. D. (2023). Beware "persuasive communication devices' when writing and reading scientific articles. *Elife*, *12*, e88654.

Cosmides, L. (1985). Deduction or Darwinian Algorithms? (Doctoral dissertation, Harvard University Cambridge, Cambridge).

Cosmides, L., J. Tooby. (1987). From Evolution to Behavior: Evolutionary Psychology as the Missing Link, in J. Dupre (ed.), *The Latest on the Best* (pp. 277-306). MIT press, Cambridge.

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II: Case study: A computational theory of social exchange. *Ethology and sociobiology*, *10*(1-3), 51-97.

Cosmides, L., J. Tooby. (1992). Cognitive adaptations for social exchange. In J.H., Barkow, L. Cosmides & J. Tooby (eds.), *The Adapted Mind* (pp. 163-228). Oxford: Oxford University Press.

Cosmides, L., & Tooby, J. (2016). Neurocognitive adaptations designed for social exchange. In D.M. Buss (ed.), *The handbook of evolutionary psychology.* Hoboken: John Wiley & Sons.

Cosmides, L., Tooby, J., & Barkow, J. H. (1992). Introduction: Evolutionary psychology and conceptual integration. *The adapted mind*, 3-15.

Currie, A. (2015). Marsupial lions and methodological omnivory: Function, success and reconstruction in paleobiology. Biology & Philosophy, 30, 187–209.

Currie, A. (2016). Ethnographic analogy, the comparative method, and archaeological special pleading. Studies in History and Philosophy of Science Part A, 55, 84-94.

Currie, A. (2018). Rock, bone, and ruin: An optimist's guide to the historical sciences. MIT Press.

Currie, A. (2021). Stepping forwards by looking back: underdetermination, epistemic scarcity and legacy data. Perspectives on Science, 29(1), 104-132.

Currie, A., & Killin, A. (2019). From things to thinking: Cognitive archaeology. Mind & Language, 34(2), 263-279.

Currie, A., Killin, A., Lequin, M., Meneganzin, A., & Pain, R. (2024). Past Materials, Past Minds: The Philosophy of Cognitive Paleoanthropology. *Philosophy Compass*.

Currie, A., & Meneganzin, A. (2022). Not by demography alone: Neanderthal extinction and null hypotheses in paleoanthropological explanation. *Biology & Philosophy*, *37*(6), 50.

Currie, A., & Sterelny, K. (2017). In defence of story-telling. *Studies in History and Philosophy of Science Part A*, *62*, 14-21.

d'Errico, F. (2003). The invisible frontier. A multiple species model for the origin of behavioral modernity. Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 12(4), 188-202.

d'Errico, F., & Colagè, I. (2018). Cultural exaptation and cultural neural reuse: A mechanism for the emergence of modern culture and behavior. *Biological Theory*, *13*(4), 213-227.

d'Errico, F., Backwell, L., 2016. Earliest evidence of personal ornaments associated with burial: The Conus shells from Border Cave. *Journal of Human Evolution* 93, 91–108, http://dx.doi.org/10.1016/j.jhevol.2016.01.002.

De Cruz, H., & De Smedt, J. (2013). The value of epistemic disagreement in scientific practice. The case of Homo floresiensis. *Studies in History and Philosophy of Science Part A*, *44*(2), 169-177.

Delisle, R. G. (2012). The disciplinary and epistemological structure of paleoanthropology: one hundred and fifty years of development. *History and philosophy of the life sciences*, 283-329.

Delisle, R. (2018). The Deceptive Search for "Missing Links" in Human Evolution, 1860-2010: Do Paleoanthropologists Always Work in the Best Interest of Their Discipline. *Rethinking Human Evolution. MIT Press, Cambridge, Massachusetts*, 1-30.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. Behavioral and Brain Sciences, 6(3), 343–355. doi: 10.1017/S0140525X00016393

Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Shuster.

De Waal, F. B. (1999). Anthropomorphism and anthropodenial: consistency in our thinking about humans and other animals. *Philosophical topics*, *27*(1), 255-280.

Dibble, H. L., Aldeias, V., Goldberg, P., McPherron, S. P., Sandgathe, D., & Steele, T. E. (2015). A critical look at evidence from La Chapelle-aux-Saints supporting an intentional Neandertal burial. Journal of Archaeological Science, 53, 649-657.

Dirks, P. H., Berger, L. R., Roberts, E. M., Kramers, J. D., Hawks, J., Randolph-Quinney, P. S., ... & Tucker, S. (2015). Geological and taphonomic context for the new hominin species Homo naledi from the Dinaledi Chamber, South Africa. *Elife*, *4*, e09561.

Dirks, P. H. G. M., Berger, L. R., Hawks, J., Randolph-Quinney, P. S., Backwell, L. R., & Roberts, E. M. (2016). Comment on "Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa?"[J. Hum. Evol. 96 (2016) 145–148]. *Journal of Human Evolution*, *96*, 149-153.

Dirks, P. H., Roberts, E. M., Hilbert-Wolf, H., Kramers, J. D., Hawks, J., Dosseto, A., ... & Berger, L. R. (2017). The age of Homo naledi and associated sediments in the Rising Star Cave, South Africa. *Elife*, *6*, e24231.

Domínguez-Rodrigo, M. (2014). Is the "savanna hypothesis" a dead concept for explaining the emergence of the earliest hominins?. *Current Anthropology*, *55*(1), 59-81.

Domínguez-Rodrigo, M., & Alcalá, L. (2016). 3.3-million-year-old stone tools and butchery traces? More evidence needed. PaleoAnthropology, 2016, 46-53.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of science*, 67(4), 559-579.

Douglas, H. E. (2009). Science, policy, and the value-free ideal. University of Pittsburgh Press.

Douven, I. (Summer 2025 Edition). Abduction. In Zalta, N. & Uri Nodelman, E. (Eds.), *The Stanford Encyclopedia of Philosophy*. URL = <https://plato.stanford.edu/archives/sum2025/entries/abduction/>.

Downes, S. M. (2021). The role of ancient DNA research in archaeology. *Topoi*, *40*(1), 285-293.

Dray, W. H. (1968). On explaining how-possibly. *The Monist, 52*(3), 390–407.

Duday, A. in Social Archaeology of Funerary Remains (eds Gowland, R. L. & Knüsel, C. J.) 30–56 (Oxbow Books, 2006).

Dutton, D. (2003). Aesthetics and evolutionary psychology. In J. Levison (Ed.), *The Oxford handbook for aesthetics*. New York: Oxford University Press.

Egeland, C. P., Domínguez-Rodrigo, M., Pickering, T. R., Menter, C. G., & Heaton, J. L. (2018). Hominin skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene. *Proceedings of the National Academy of Sciences*, *115*(18), 4601-4606.

Einwögerer, T., Friesinger, H., Händel, M., Neugebauer-Maresch, C., Simon, U., & Teschler-Nicola, M. (2006). Upper Palaeolithic infant burials. *Nature*, *444*(7117), 285-285.

eLife, 2022. eLife's new model: changing the way you share your research | Inside eLife | eLife [WWW Document]. https://elifesciences.org/inside-elife/54d63486/elife-s-new-model-changing-the-way-you-share-yourresearch

Elliott, M., MAKHUBELA, T., Brophy, J., Churchill, S., Peixotto, B., FEUERRIEGEL, E., ... & BERGER, L. (2021). Expanded explorations of the Dinaledi subsystem, Rising Star cave system, South Africa. *PaleoAnthropology*, *2021*(1), 15-22.

Falk, J. H., & Balling, J. D. (2010). Evolutionary influence on human landscape preference. *Environment and behavior*, *42*(4), 479-493.

Fessler, D.M.T., Clark, J.A., & Clint, E.K. (2016). Evolutionary psychology and evolutionary anthropology. In D.M. Buss (Ed.), *The handbook of evolutionary psychology: Integrations* (2nd ed., pp. 1029-1046). John Wiley & Sons, Inc..

Fisher, R. A. (1935). *The design of experiments*.

Foister, T.I., Žliobaitė, I., Wilson, O.E., Fortelius, M., & Tallavaara, M. (2023). Homo heterogenus: Variability in early Pleistocene Homo environments. *Evolutionary Anthropology: Issues, News, and Reviews*, *32*(6), 373-385.

Foecke, K. K., Queffelec, A., Pickering, R., & Harvati, K. (2024). No Sedimentological Evidence for Deliberate Burial by Homo naledi–A Case Study Highlighting the Need for Best Practices in Geochemical Studies Within Archaeology and Paleoanthropology. *Paleoanthropology*.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Fogelin, L. (2007). Inference to the best explanation: A common and effective form of archaeological reasoning. *American antiquity*, *72*(4), 603-626.

Forber, P. (2011). Reconceiving eliminative inference. Philosophy of Science, 78(2), 185-208.

Forber, P., & Griffith, E. (2011). Historical reconstruction: Gaining access to the deep past. *Philosophy and Theory in Biology*, 3, 1-19.

Formicola, V. (2007). From the Sunghir children to the Romito dwarf: aspects of the Upper Paleolithic funerary landscape. Current Anthropology, 48(3), 446-453.

Fotiadis, Michael. 2018. Knowledge and Inference. In *The Encyclopedia of Archaeological Sciences*, edited by Sandra L. López Varela, pp. 1‑5. Wiley-Blackwell, Hoboken, New Jersey.

Fu, Q., Cao, P., Dai, Q., Bennett, E. A., Feng, X., Yang, M. A., ... & Ji, Q. (2025). Denisovan mitochondrial DNA from dental calculus of the> 146,000-year-old Harbin cranium. *Cell*.

Fuentes Agustin, Kissel Marc, Spikins Penny, Molopyane Keneiloe, Hawks John, Berger Lee R. (2023) Burials and engravings in a small-brained hominin, Homo naledi, from the late Pleistocene: contexts and evolutionary implications eLife 12:RP89125. https://doi.org/10.7554/eLife.89125.1

Gargett, R.H. (1989). Grave Shortcomings: The Evidence for Neanderthal Burial. *Current Anthropology*, 30(2), pp. 157–190.

Gargett, R. H. (1999). Middle Palaeolithic burial is not a dead issue: the view from Qafzeh, Saint-Césaire, Kebara, Amud, and Dederiyeh. *Journal of human evolution*, *37*(1), 27-90.

Garvey, R. (2018). Current and potential roles of archaeology in the development of cultural evolutionary theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170057.

Garvin Heather M., Elliott Marina C., Delezene Lucas K., Hawks John, Churchill Steven E., Berger Lee R., Holliday Trenton W. (2017) Body size, brain size, and sexual dimorphism in Homo naledi from the Dinaledi Chamber *Journal of Human Evolution* **111**:119‑138

Gee, H. (2000), *In Search of Deep Time*. New York: The Free Press.

Ghasidian, E., Kafash, A., Kehl, M., Yousefi, M., Heydari-Guran, S., 2023. Modelling Neanderthals' dispersal routes from Caucasus towards east. PLoS One 18 (2) , http://dx.doi.org/10.1371/journal.pone.0281978 (article no e0281978).

Gonçalves, A., & Carvalho, S. (2019). Death among primates: a critical review of non-human primate interactions towards their dead and dying. *Biological Reviews*, 94(*4*), 1502-1529.

Gould, S.J., & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme, *Proc. R. Soc. Lond*. B.205581-598, http://doi.org/10.1098/rspb.1979.0086.

Gray, R.D., Heaney, M., & Fairhall, S. (2003). Evolutionary psychology and the challenge of adaptive explanation. In K. Sterelny, & J. Fitness (eds.), *From mating to mentality: Evaluating evolutionary psychology* (pp. 247-268). Londra: Routledge.

Griffiths, P.E., (2008). Ethology, Sociobiology and Evolutionary Psychology. In S. Sarkar, & A. Plutynski (eds.), *Blackwell's Companion to Philosophy of Biology* (pp. 393-414). Oxford: Blackwells

Halina, M. (2024). *Animal Minds*. Cambridge University Press.

Hall, S., & Binneman, J. (1987). Later Stone Age burial variability in the Cape: a social interpretation. *The South African Archaeological Bulletin*, 140-152.

Hammond, M., & Stocking, G. W. (1988). The shadow man paradigm in palaeoanthropology: 1911–1945. *Bones. Bodies. Behavior: Essavs on Biological Anthropology. George W. Stockinged.(Madison: University of Wisconsin Press, 1988)*, 117-137.

Han, K.T. (2007). Responses to six major terrestrial biomes in terms of scenic beauty, preference, and restorativeness. *Environment and Behavior*, *39*(4), 529-556.

Hanen, M., and Jane, K. (1989). Inference to the Best Explanation in Archaeology. In *Critical Traditions in Contemporary Archaeology: Essays in the Philosophy, History, and Socio-Politics of Archaeology*, edited by Valerie Pinsky and Alison Wylie, pp. 14‑17. Cambridge University Press, Cambridge

Harman, Gilbert H. 1965. The Inference to the Best Explanation. Philosophical Review 74(1):88‑95.

Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., Boës, X., Quinn, R. L., Brenet, M., Arroyo, A., Taylor, N., Clément, S., Daver, G., Brugal, J. P., Leakey, L., Mortlock, R. A., Wright, J. D., Lokorodi, S., Kirwa, C., Kent, D. V., … Roche, H. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. Nature, 521(7552), 310–315. https://doi.org/10.1038/nature14464

Hartmann, P., & Apaolaza-Ibáñez, V. (2010). Beyond savanna: An evolutionary and environmental psychology approach to behavioral effects of nature scenery in green advertising. *Journal of Environmental Psychology*, *30*(1), 119-128.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve?. *science*, *298*(5598), 1569-1579.

Havstad, J. C. (2022). Sensational science, archaic hominin genetics, and amplified inductive risk. *Canadian Journal of Philosophy*, *52*(3), 295-320.

Hawks, J. (2014). Still evolving (after all these years). *Scientific American*, *311*(3), 86-91.

Hawks, J., Wang, E.T., Cochran, G.M., Harpending, H.C., & Moyzis, R.K. (2007). Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences*, *104*(52), 20753-20758.

Hawks, J., Elliott, M., Schmid, P., Churchill, S. E., Ruiter, D. J. de, Roberts, E. M., Hilbert-Wolf, H., Garvin, H. M., Williams, S. A., Delezene, L. K., Feuerriegel, E. M., Randolph-Quinney, P., Kivell, T. L., Laird, M. F., Tawane, G., DeSilva, J. M., Bailey, S. E., Brophy, J. K., Meyer, M. R., … Berger, L. R. (2017). New fossil remains of Homo naledi from the Lesedi Chamber, South Africa. eLife, 6, e24232. https://doi.org/10.7554/eLife.24232

Hempel, C. G. (1942). The function of general laws in history. *Journal of Philosophy*, 39(2), 35-48.

Henke, W. (2015). Historical overview of paleoanthropological research. In *Handbook of paleoanthropology* (pp. 3-95). Springer, Berlin, Heidelberg.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*(2-3), 61-83.

Henshilwood, C.S., *et al.* (2002). *Emergence of modern human behavior: Middle Stone Age engravings from South Africa*. Science, 295(5558), 1278-1280.

Henshilwood, C. S., & Marean, C. W. (2003). The origin of modern human behavior: Critique of the models and their test implications. Current Anthropology, 44, 627–651.

Herbert, T. D. (2023). The mid-Pleistocene climate transition. *Annual Review of Earth and Planetary Sciences*, *51*(1), 389-418.

Hershkovitz, I., Weber, G. W., Quam, R., Duval, M., Grün, R., Kinsley, L., ... & Weinstein-Evron, M. (2018). The earliest modern humans outside Africa. *Science*, *359*(6374), 456-459.

Hoffmann, D. L., Standish, C. D., Pike, A. W., García-Diez, M., Pettitt, P. B., Angelucci, D. E., ... & Zilhão, J. (2018). Dates for Neanderthal art and symbolic behaviour are reliable. *Nature Ecology & Evolution*, *2*(7), 1044-1045.

Holliday, T. W. (1997). Body proportions in Late Pleistocene Europe and modern human origins. *Journal of human evolution*, *32*(5), 423-448.

Holloway Ralph L., Hurst Shawn D., Garvin Heather M., Thomas Schoenemann P., Vanti William B., Berger Lee R., Hawks John (2018) Endocast Morphology of Homo Naledi from the Dinaledi Chamber, South Africa Proceedings of the National Academy of Sciences of the United States of America 115:5738–43 https://doi.org/10.1073/pnas.1720842115

Horsburgh, K. A. (2015). Molecular anthropology: The judicial use of genetic data in archaeology. *Journal of archaeological science*, *56*, 141-145.

Hovers, E., Rak, Y., Lavi, R., & Kimbel, W. H. (1995). Hominid remains from Amud Cave in the context of the Levantine Middle Paleolithic. Paléorient, 47-61.

Hovers, Erella, and Anna Belfer-Cohen, ' Insights into Early Mortuary Practices of *Homo*', in Liv Nilsson Stutz, and Sarah Tarlow (eds), *The Oxford Handbook of the Archaeology of Death and Burial* (2013; online edn, Oxford Academic, 1 Aug. 2013), https://doi.org/10.1093/oxfordhb/9780199569069.013.0035

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., ... & Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194-197.

Hurst S. D., Holloway R. L., Balzeau A., Garvin H. M., Vanti W. B., Berger L. R., Hawks J (2024) The endocast morphology of LES1, Homo naledi. American Journal of Biological Anthropology *e* 24983 https://doi.org/10.1002/ajpa.24983

Jacoby, F. (2023). Evidence, justification, and epistemic standards. *Synthese*, *201*(2), 44.

Jones, E. D. (2019). Ancient genetics to ancient genomics: celebrity and credibility in data-driven practice. Biology & Philosophy, 34(2), 27.

Jeffares, B. (2008). Testing times: regularities in the historical sciences. Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences, 39(4), 469-475. Joordens, J. C., d'Errico, F., Wesselingh, F. P., Munro, S., De Vos, J., Wallinga, J., ... & Roebroeks, W. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. Nature, 518(7538), 228-231.

Kelly, D. R. (2011). *Yuck! The nature and moral significance of disgust*. Cambridge, MA: MIT Press.

Killin, A. (2018). The origins of music: Evidence, theory, and prospects. *Music & Science*, *1*, 2059204317751971.

Killin, A., & Pain, R. (2023). Cognitive archaeology and the minimum necessary competence problem. Biological Theory, 18(4), 269-283.

Kissel, M., & Fuentes, A. (2018). "Behavioral modernity'as a process, not an event, in the human niche. Time and Mind, 11(2), 163-183.

Kissel, M., & Fuentes, A. (2021). The ripples of modernity: How we can extend paleoanthropology with the extended evolutionary synthesis. Evolutionary Anthropology: Issues, News, and Reviews, 30(1), 84-98.

Kitcher, P. (1985). *Vaulting Ambition. Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.

Klein, R. G. (2008). Out of Africa and the evolution of human behavior. Evolutionary Anthropology, 17(6):267–281.

Klima, B. (1987). A triple burial from the Upper Paleolithic of Dolní Věstonice, Czechoslovakia. *Journal of Human Evolution*, *16*(7-8), 831-835.

Kokkov, K. (2019). Warrants, middle-range theories, and inferential scaffolding in archaeological interpretation. Perspectives on Science, 27(2), 171-186.

Lahr, M., & Foley, R. (2004). Human evolution writ small. *Nature*, *431*(7012), 1043-1044.

Landau, M. (1991). *Narratives in human evolution*. Yale University Press: New Haven.

Laland, K.N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics*, *11*(2), 137-148.

Laland, K.N., & Brown, G.R. (2011). *Sense and nonsense: Evolutionary perspectives on human behaviour*. Oxford: Oxford University Press.

Leonelli, S. (2015). What counts as scientific data? A relational framework. Philosophy of Science, 82(5), 810-821.

Leonelli, S. (2017). Chapman and Wylie's Evidential Reasoning in Archaeology. BJPS Review of Books.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. University of Chicago Press.

Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. Science, Technology, & Human Values, 42(2), 191-202.

Lequin, M. (2018). Paleoanthropology's uses of the bipedal criterion. History and Philosophy of the Life Sciences, 40(1), 7.

Lequin, M., Colard, T., Colombo, A., Le Cabec, A., Remy, F., & Schuh, A. (2025). Investigating development in human evolution: specificities, challenges, and opportunities. Evolutionary Anthropology: Issues, News, and Reviews, 34(1), e70001.

Levin, N. E. (2015). Environment and climate of early human evolution. *Annual Review of Earth and Planetary Sciences*, *43*(1), 405-429.

Lewontin, R. (1998). The evolution of cognition: Questions we will never answer. *Methods, models and conceptual issues (An invitation to cognitive science 4)*, ed. by Don Scarborough and Saul Stern berg, 108-32.

Lieberman, D., Tooby, J. & Cosmides, L. (2007). The architecture of human kin detection. *Nature* 445, 727-731. https://doi.org/10.1038/nature05510

Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. Routledge, London.

Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*, *20*(1).

Liu, H; Prugnolle, F; Manica, A; Balloux, F. (2006). "A geographically explicit genetic model of worldwide human-settlement history". *American Journal of Human Genetics*. **79** (2): 230–237.

Lloyd, E.A. (1999). Evolutionary psychology: The burdens of proof. *Biology and Philosophy*, 14(2), 211-233.

Lloyd, E. A. (2015). Adaptationism and the logic of research questions: how to think clearly about evolutionary causes. *Biological Theory*, 10(4), 343-362.

Lloyd, E. (2021). *Adaptation*. Cambridge: Cambridge University Press.

Lobue, V., & DeLoache, J.S. (2008). Detecting the snake in the grass: attention

Love, A. C. (2008). Explaining evolutionary innovations and novelties: Criteria of explanatory adequacy and epistemological prerequisites. *Philosophy of science*, *75*(5), 874-886.

Macphail, E. M. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences*, *10*(4), 645-656.

Makedonska, J. A., & Strait, D. S. (2015). Hypothesis Testing in Palaeoanthropology. *eLS*, 1-5.

Marginedas, F., Saladié, P., Połtowicz-Bobak, M., Terberger, T., Bobak, D., & Rodríguez-Hidalgo, A. (2025). New insights of cultural cannibalism amongst Magdalenian groups at Maszycka Cave, Poland. *Scientific Reports*, *15*(1), 2351.

Marrone R (1998) Grieving and mourning: distinctions in process. Illn Crisis Loss 6(3):320–332

Martinón-Torres, M., d'Errico, F., Santos, E., Álvaro Gallo, A., Amano, N., Archer, W., ... & Petraglia, M. D. (2021). Earliest known human burial in Africa. *Nature*, *593*(7857), 95-100.

Martinón-Torres, M., Garate, D., Herries, A. I., & Petraglia, M. D. (2023). No scientific evidence that Homo naledi buried their dead and produced rock art. *Journal of Human Evolution*, 103464.

Maureille, B., & Knüsel, C. J. (2022). The earliest European burials. In *The Routledge Handbook of Archaeothanatology* (pp. 140-158). Routledge.

McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of human evolution*, *39*(5), 453-563.

McGrew, W. C., Falótico, T., Gumert, M. D., & Ottoni, E. B. (2019). A simian view of the Oldowan: Reconstructing the evolutionary origins of human technology. In K. A. Overmann & F. L. Coolidge (Eds.), Squeezing minds from stones: Cognitive archaeology & the evolution of the human mind (pp. 13–41). Oxford University Press. https://doi.org/10.1093/oso/9780190854614.003.0002

Mekel-Bobrov, N., Gilbert, S.L., Evans, P.D., Vallender, E.J., Anderson, J.R., Hudson, R.R., … & Lahn, B.T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science*, 309(5741), 1720-1722.

Meketa, Irina, 2014, "A Critique of the Principle of Cognitive Simplicity in Comparative Cognition", Biology & Philosophy, 29(5): 731–745. doi:10.1007/s10539-014-9429-z

Mellars, P. (2010). Neanderthal symbolism and ornament manufacture: The bursting of a bubble? Proceedings of the National Academy of Sciences of the United States of America, 107(47):20147–20148.

Meneganzin, A., & Bernardi, M. (2023). Were Neanderthals and Homo sapiens "good species'?. *Quaternary Science Reviews*, *303*, 107975.

Meneganzin, A., & Currie, A. (2022). Behavioural modernity, investigative disintegration & Rubicon expectation. *Synthese*, *200*(1), 47.

Meneganzin, A., & Currie, A. (2025). Not wasted on the young: Childhood, trait complexes & human behavioral ecology. *Studies in History and Philosophy of Science*, *109*, 12-20.

Meneganzin, A., & Killin, A. (2024). Beyond Reasonable Doubt: Reconsidering Neanderthal Aesthetic Capacity. *Phenomenology and the Cognitive Sciences*.

Meneganzin, A., Pievani, T., & Manzi, G. (2022). Pan-Africanism vs. single-origin of Homo sapiens: Putting the debate in the light of evolutionary biology. *Evolutionary Anthropology: Issues, News, and Reviews*, *31*(4), 199-212.

Meneganzin, A., Ramsey, G., & DiFrisco, J. (2024). What is a trait? Lessons from the human chin. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *342*(2), 65-75.

Meneganzin, A., & Stringer, C. (2024). Homo sapiens, Neanderthals and speciation complexity in palaeoanthropology. *Evolutionary Journal of the Linnean Society*, *3*(1), kzae033.

Mercader, J., Akuku, P., Boivin, N., Camacho, A., Carter, T., Clarke, S., ... & Durkin, P. (2025). Homo erectus adapted to steppe-desert climate extremes one million years ago. Communications Earth & Environment, 6(1), 1.

Mikhalevich, I. (2015). Experiment and Animal Minds: Why the Choice of the Null Hypothesis Matters. Philosophy of Science, 82(5), 1059–1069. doi: 10.1086/683440

Miller, G.F. (2000). *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*.

Mithen, S. (1996). The prehistory of the mind: a search for the origins of art, religion and science. London: Thames and Hudson.

Mithen, S. (2014). The cognition of Homo neanderthalensis and H. sapiens: Does the use of pigment necessarily imply symbolic thought? In Dynamics of Learning in Neanderthals and Modern Humans Volume 2, pages 7–16. Springer.

Mithen, S. (2024). *The language puzzle: How we talked our way out of the stone age*. London: Profile Books.

Molari, C. 1994. The industry on bone of the Pleistocene layers from the Arene Candide Cave (Savona, Italy). *Quaternaria Nova* IV, 297–340.

Monsó, S. (2024). Playing possum: How animals understand death. In *Playing Possum*. Princeton University Press.

Montgomery, S. H., & Mundy, N. I. (2013). Parallel episodes of phyletic dwarfism in callitrichid and cheirogaleid primates. *Journal of Evolutionary Biology*, *26*(4), 810-819. https://doi.org/10.1111/jeb.12097

Morgan, C. L. (1903). An introduction to comparative psychology. (New ed. Revised). London: The Walter Scott Publishing Co.

Moura, J.M., Ferreira Júnior, W.S., Silva, T.C., & Albuquerque, U.P. (2018). The influence of the evolutionary past on the mind: An analysis of the preference for landscapes in the human species. *Frontiers in Psychology*, *9*, 2485.

Nelson, L.H. (2017). *Biology and feminism: A philosophical introduction*. Cambridge University Press.

Neyman, J., & Pearson, E. S. (1933, October). The testing of statistical hypotheses in relation to probabilities a priori. In Mathematical proceedings of the Cambridge philosophical society (Vol. 29, No. 4, pp. 492-510). Cambridge University Press.

Niekus, M. J. T., Kozowyk, P. R., Langejans, G. H., Ngan-Tillard, D., van Keulen, H., van der Plicht, J., ... & Dusseldorp, G. L. (2019). Middle Paleolithic complex technology and a Neandertal tar-backed tool from the Dutch North Sea. Proceedings of the National Academy of Sciences, 116(44), 22081-22087.

Norton, J. D. (2021). *The material theory of induction* (p. 680). University of Calgary Press.

Nowell, A. (2023). Rethinking neandertals. *Annual Review of Anthropology*, *52*(1), 151-170

Öhman, A., & Mineka, S. (2001). Fear, phobias and preparedness: Toward an evolved module of fear and fear learning. Psychological Review, 108, 483-522.

Oktaviana, A. A., Joannes-Boyau, R., Hakim, B., Burhan, B., Sardi, R., Adhityatama, S., ... & Aubert, M. (2024). Narrative cave art in Indonesia by 51,200 years ago. *Nature*, *631*(8022), 814-818.

Orians, G. (1980). Habitat selection: General theory and applications to human behavior. In J.S. Lockard (Ed.), *The evolution of human social behavior* (pp. 49-66). Chicago: Elsevier.

Orians, G. (1986). An ecological and evolutionary approach to landscape aesthetics. In E.C. Penning-Rowsell & D. Lowenthal (Eds.), *Landscape meaning and values* (pp. 3-25). London: Allen & Unwin.

Orians, G., & Heerwagen, J.H. (1992). Evolved responses to landscapes. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: University Press.

Orschiedt, J. (2002) Secondary burial in the Magdalenian: the Brillenhöhle (Blauberen, Southwest Germany). *Paleo* 14: 241–56.

Orschiedt, J. (2018). The Late Upper Palaeolithic and earliest Mesolithic evidence of burials in Europe. *Philosophical transactions of the Royal Society B: Biological sciences*, 373(1754), 20170264.

Pain, R. (2019). What Can the Lithic Record Tell Us About the Evolution of Hominin Cognition? Topoi 40 (1):245-259.

Parravicini, A., & Pievani, T. (2016). Multi-level human evolution: ecological patterns in hominin phylogeny. Journal of Anthropological Sciences, 94, 167-182.

Parravicini, A., & Pievani, T. (2018). Continuity and discontinuity in human language evolution: putting an old-fashioned debate in its historical perspective. Topoi, 37(2), 279-287.

Parravicini, A., & Pievani, T. (2019). Mosaic evolution in hominin phylogeny: meanings, implications, and explanations. Journal of Anthropological Sciences, 97, 45-68.

Pearson, M. P. (1999). *The archaeology of death and burial*. Stroud: Sutton.

Peeters, S., & Zwart, H. (2020). Neanderthals as familiar strangers and the human spark: How the "golden years' of Neanderthal research reopen the question of human uniqueness. History and philosophy of the life sciences, 42(3), 33.

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 731–744

Peresani, M., Fiore, I., Gala, M., Romandini, M., & Tagliacozzo, A. (2011). Late Neandertals and the intentional removal of feathers as evidenced from bird bone taphonomy at Fumane Cave 44 ky BP, Italy. *Proceedings of the National Academy of Sciences*, *108*(10), 3888-3893.

Pettitt, P. (2002). The Neanderthal dead: exploring mortuary variability in Middle Palaeolithic Eurasia. *Before farming*, *2002*(1), 1-26.

Pettitt, P., (2011). The Palaeolithic Origins of Human Burial. Routledge, London.

Pettitt, P. (2018). Hominin evolutionary thanatology from the mortuary to funerary realm: the palaeoanthropological bridge between chemistry and culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1754), 20180212.

Pettitt P. Did *Homo naledi* dispose of their dead in the Rising Star Cave system? S Afr J Sci. 2022;118(11/12), Art. #15140. https://doi.org/10.17159/sajs.2022/15140

Pettitt, P. B., Richards, M., Maggi, R., & Formicola, V. (2003). The Gravettian burial known as the Prince ("Il Principe"): new evidence for his age and diet. *Antiquity*, *77*(295), 15-19.

Pettitt, P., & Wood, B. (2024). What we know and do not know after the first decade of Homo naledi. *Nature ecology & evolution*, 8(9), 1579–1583. https://doi.org/10.1038/s41559-024-02470-0

Pickering, R., & Kgotleng, D. W. (2024). Preprints, press releases and fossils in space: What is happening in South African human evolution research?. *South African Journal of Science*, *120*(3/4). https://doi.org/10.17159/sajs.2024/17473

Pickford, M., & Senut, B. (2001). The geological and faunal context of Late Miocene hominid remains from Lukeino, Kenya. *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, *332*(2), 145-152.

Pinker, S. (1997). *How the Mind Works*. London: Penguin Books.

Pomeroy, E., Bennett, P., Hunt, C. O., Reynolds, T., Farr, L., Frouin, M., Holman, J., Lane, R., French, C., & Barker, G. (2020a). New Neanderthal remains associated with the "flower burial' at Shanidar Cave. *Antiquity*, *94*(373), 11–26. https://doi.org/10.15184/aqy.2019.207

Pomeroy, E., Hunt, C. O., Reynolds, T., Abdulmutalb, D., Asouti, E., Bennett, P., Bosch, M., Burke, A., Farr, L., Foley, R., French, C., Frumkin, A., Goldberg, P., Hill, E., Kabukcu, C., Lahr, M. M., Lane, R., Marean, C., Maureille, B., ⋯ Barker, G. (2020b). Issues of theory and method in the analysis of Paleolithic mortuary behavior: A view from Shanidar Cave. *Evolutionary Anthropology: Issues, News, and Reviews*, *29*(5), 263–279. https://doi.org/10.1002/evan.21854

Potts, R. (1998). Environmental hypotheses of hominin evolution. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *107*(S27), 93-136.

Potts, R. (2012). Evolution and environmental change in early human prehistory. *Annual Review of Anthropology*, *41*(1), 151-167.

Powell, R. (2012). The Future of Human Evolution. *The British Journal for the Philosophy of Science*, *63*(1), 145-175. https://doi.org/10.1093/bjps/axr027

Premo, L. S., & Kuhn, S. L. (2010). Modeling effects of local extinctions on culture change and diversity in the Paleolithic. PloS one, 5(12), e15582.

Rathmann, J., Korpela, K.M., & Stojakowits, P. (2022). Pleistocene hypothesis–moving savanna perceptual preference hypothesis beyond savanna. *Frontiers in Psychology*, *13*, 901799.

Reiss, J. (2015). A pragmatist theory of evidence. Philosophy of Science, 82(3), 341-362.

Rendu, W., Beauval, C., Crevecoeur, I., Bayle, P., Balzeau, A., Bismuth, T., Bourguignon, L., Delfour, G., Faivre, J. P., Lacrampe-Cuyaubère, F., Tavormina, C., Todisco, D., Turq, A., & Maureille, B. (2014). Evidence supporting an intentional Neandertal burial at La Chapelle-aux-Saints. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 81–86. https://doi.org/10.1073/pnas.1316780110

Perreault, C. (2019). *The Quality of the Archaeological Record*. University of Chicago Press, Chicago.

Pickering, T. R. (2013). *Rough and tumble: aggression, hunting, and human evolution*. Univ of California Press.

Pievani, T. (2014). *Evoluti e abbandonati. Sesso, politica, morale: Darwin spiega proprio tutto?* Torino: Einaudi.

Profico, A., Di Vincenzo, F., Lorenza, G., Marcello, P., & Manzi, G. (2016). Filling the gap. Human cranial remains from Gombore II (Melka Kunture, Ethiopia; ca. 850 ka) and the origin of Homo heidelbergensis. *Journal of Anthropological Sciences*, *94*, 41-63.

Renfrew, C. (1996). The sapient behaviour paradox: How to test for potential? In P. Mellars & K. Gibson (Eds.), Modelling the early human mind (pp. 11–15). McDonald Institute.

Richardson, R.C. (2007). Evolutionary psychology as maladapted psychology. Manchester: MIT press.

Riel-Salvatore J., Clark, G. A. (2001). Middle and Early Upper Paleolithic Burials and the Use of Chronotypology in Contemporary Paleolithic Research. *Current Anthropology*, *42*(4), 449-479.

Rightmire, G. P. (2001). Patterns of hominid evolution and dispersal in the Middle Pleistocene. *Quaternary International*, *75*(1), 77-84.

Roberts, P., Boivin, N., Lee-Thorp, J., Petraglia, M., & Stock, J. (2016). Tropical forests and the genus Homo. *Evolutionary Anthropology: Issues, News, and Reviews*, *25*(6), 306-317.

Roberts, P., & Stewart, B.A. (2018). Defining the "generalist specialist" niche for Pleistocene Homo sapiens. *Nature Human Behaviour*, *2*(8), 542-550.

Roberts, P., & Petraglia, M. (2015). Pleistocene rainforests: barriers or attractive environments for early human foragers?. *World Archaeology*, *47*(5), 718-739.

Roebroeks, W., Sier, M. J., Nielsen, T. K., De Loecker, D., Parés, J. M., Arps, C. E., & Mücher, H. J. (2012). Use of red ochre by early Neandertals. *Proceedings of the National Academy of Sciences*, *109*(6), 1889-1894.

Rossano, M. J. (2015). The evolutionary emergence of costly rituals. *PaleoAnthropology*, 2015, 78-100.

Russell, M. (1987) Mortuary practice at the Krapina Neanderthal site. *American Journal of Physical Anthropology* 72: 381–97.

Sandgathe, D. M., Dibble, H. L., Goldberg, P., & McPherron, S. P. (2011). The Roc de Marsal Neandertal child: A reassessment of its status as a deliberate burial. Journal of human evolution, 61(3), 243-253.

Scerri, E.M., Roberts, P., Yoshi Maezumi, S., & Malhi, Y. (2022). Tropical forests in the deep human past. *Philosophical Transactions of the Royal Society B*, *377*(1849), 20200500

Scerri, E. M., & Will, M. (2023). The revolution that still isn't: The origins of behavioral complexity in Homo sapiens. Journal of Human Evolution, 179, 103358.

Schwartz, J. H. (2018). What's real about human evolution? Received wisdom, assumptions, and scenarios. In Schwartz, J. H. (Ed.). *Rethinking Human Evolution. MIT Press, Cambridge, MA*, 61-91.

Senut, B., & Pickford, M. (2001). 'Millennium Ancestor', a 6-million-year-old bipedal hominid from Kenya: news & views. *South African Journal of Science*, *97*(1), 22.

Sepulchre, P., Ramstein, G., Fluteau, F., Schuster, M., Tiercelin, J. J., & Brunet, M. (2006). Tectonic uplift and Eastern Africa aridification. *Science*, *313*(5792), 1419-1423.

Sergi, S., R. Parenti & G. Paoli. (1974). Il giovane paleolitico della Caverna delle Arene Candide. *Memorie dell'Istituto Italiano di Paleontologia Umana*, n.s. 2:13–38.

Sheets-Johnstone, M. (1986). On the conceptual origin of death. *Philosophy and Phenomenological Research*, 47(1), 31-58.

Shettleworth, S. J. (2010a). Clever animals and killjoy explanations in comparative psychology. Trends in cognitive sciences, 14(11), 477-481.

Shettleworth, S. J. (2010b). Cognition, evolution, and behavior (2nd edition). New York: Oxford university press.

Shipton C, Clarkson C, Cobden R (2018) "Were Acheulean bifaces deliberately made symmetrical? Archaeological and experimental evidence" Cambridge Archaeological Journal 29(1): 65-79.

Shipton, C., Roberts, P., Archer, W., Armitage, S.J., Bita, C., Blinkhorn, J., … & Boivin, N. (2018). 78,000-year-old record of Middle and Later Stone Age innovation in an East African tropical forest. *Nature Communications*, *9*(1), 1832.

Silverman, I. and Choi, J. (2015). Spatial Navigation and Landscape Preferences. In The Handbook of Evolutionary Psychology, D.M. Buss (Ed.). https://doi.org/10.1002/9781119125563.evpsych108

Smith, R. J. (2016). Explanations for adaptations, just-so stories, and limitations on evidence in evolutionary biology. *Evolutionary Anthropology: Issues, News, and Reviews*, *25*(6), 276-287.

Smith, R. J. (2018). The continuing misuse of null hypothesis significance testing in biological anthropology. American Journal of Physical Anthropology, 166(1), 236-245.

Smith, R. J., & Wood, B. (2017). The principles and practice of human evolution research: Are we asking questions that can be answered?. *Comptes Rendus Palevol*, *16*(5-6), 670-679.

Smith, R. J., & Wood, B. (2024). On the scientific credibility of paleoanthropology: Reply to Villmoare and Kimbel. *Evolutionary Anthropology: Issues, News, and Reviews*, *33*(5), e22041.

Sober, E. (2001). The principle of conservatism in cognitive ethology. Royal Institute of Philosophy Supplements, 49, 225-238.

Sober, E. (2005). "Comparative Psychology Meets Evolutionary Biology: Morgan's Canon and Cladistic Parsimony." In Thinking with Animals: New Perspectives on Anthropomorphism, ed. Lorraine Daston and Gregg Mitman, 85–99. New York: Columbia University Press.

Sommer, R. (1997). Further cross-national studies of tree form preference. *Ecological Psychology*, *9*(2), 153-160.

Sommer, J. D. (1999). The Shanidar IV 'flower burial': a re-evaluation of Neanderthal burial ritual. *Cambridge Archaeological Journal*, *9*(1), 127-129.

Stanford, P. K. (2006). Exceeding our grasp: Science, history, and the problem of unconceived alternatives. Oxford: Oxford University Press.

Sterelny, K. (2011). From hominins to humans: how sapiens became behaviourally modern. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566), 809-822.

Sterelny, K., Griffiths, P. E., & Griffiths, P. (1999). *Sex and death: An introduction to philosophy of biology*. University of Chicago press.

Sterelny, K., & Hiscock, P. (2024). Cumulative culture, archaeology, and the zone of latent solutions. *Current Anthropology*, *65*(1), 23-48.

Sterelny, K., & Hiscock, P. (2025). Material symbols and the archaeological record. Biological Theory, 1-11.

Stibbard-Hawkes, D. N. (2025a). Reconsidering the link between past material culture and cognition in light of contemporary hunter–gatherer material use. Behavioral and Brain Sciences, 48, e1.

Stibbard-Hawkes, D. N. (2025b). Hominin cognition: The null hypothesis. Behavioral and Brain Sciences, 48, e23. Author's response to the target article commentaries.

Stibbard-Hawkes, D. N. E., & Apicella, C. L. (2022). Myopia rates among Hadza huntergatherers are low but not exceptional. American Journal of Biological Anthropology, 179(4), 655–667.

Stiner, M. C. (2017). Love and death in the Stone Age: What constitutes first evidence of mortuary treatment of the human body?. *Biological Theory*, *12*, 248-261.

Stringer, C. (2016). The origin and evolution of Homo sapiens. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1698), 20150237.

Sykes, R. W. (2020). *Kindred: Neanderthal life, love, death and art.* London: Bloomsbury.

Tattersall, I., & Eldredge, N. (1977). Fact, theory, and fantasy in human paleontology. *American Scientist*, *65*(2), 204-211.

Tattersall, I. (1992). Species concepts and species identification in human evolution. *Journal of Human Evolution*, *22*(4-5), 341-349.

Taylor, A. H., Bastos, A. P., Brown, R. L., & Allen, C. (2022). The signature-testing approach to mapping biological and artificial intelligences. *Trends in Cognitive Sciences*, *26*(9), 738-750.

Tempini, N., & Leonelli, S. (2021). Actionable data for precision oncology: Framing trustworthy evidence for exploratory research and clinical diagnostics. Social Science & Medicine, 272, 113760.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1528), 2405-2415.

Tennie, C., Braun, D. R., Premo, L. S., & McPherron, S. P. (2016, January). The Island test for cumulative culture in the Paleolithic. In The nature of culture: based on an interdisciplinary symposium "the nature of culture', Tübingen, Germany (pp. 121-133). Dordrecht: Springer Netherlands.

Tennie, C., Premo, L. S., Braun, D. R., & McPherron, S. P. (2017). Early stone tools and cultural transmission: Resetting the null hypothesis. Current Anthropology, 58(5), 652-672.

Thomas, R., & Darvill, T. (2022). What haven't we found? Recognising the value of negative evidence in archaeology. *Antiquity*, *96*(388), 955-967.

Thieme, H. (1997). Lower Palaeolithic hunting spears from Germany. *Nature*, *385*(6619), 807-810.

Tillier, A.M., 2022. Early Primary Burials: Evidence from Southwestern Asia. In: Knusel, C., Schotsmans, E.M.J. (Eds.), The Routledge Handbook of Archaeothanatology. Bioarchaeology of mortuary behaviour. Routledge, London, pp. 124–139.

Tomasello, M. 1990 Cultural transmission in the tool use and communicatory signaling of chimpanzees? In "Language' and intelligence in monkeys and apes: comparative developmental perspectives (eds S. T. Parker & K. R. Gibson), pp. 274–311. New York, NY: Cambridge University Press.

Tomasello, M. 1996 Do apes ape? In Social learning in animals: the roots of culture (eds C. M. Heyes & B. G. Galef), pp. 319–346. San Diego, CA: Academic Press, Inc.

Tomasello, M. 1999 The cultural origins of human cognition. Cambridge, MA: Harvard University Press.

Tomasello, M. Social cognition and metacognition in great apes: a theory. Anim Cogn 26, 25–35 (2023). https://doi.org/10.1007/s10071-022-01662-0

Tomasello, M. & Call, J. 1997 Primate cognition. New York, NY: Oxford University Press.

Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, part I: Theoretical considerations. *Ethology and sociobiology*, *10*(1-3), 29-49.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, 1. Cosmides, & J. Tooby (eds.), *The Adapted Mind* (pp. 19-136). Oxford: Oxford University Press.

Tooby, J., Cosmides, L. (2016). The Theoretical Foundations of Evolutionary Psychology. In D.M. Buss (ed) *The Handbook of Evolutionary Psychology, 2nd edition* (pp. 3-87). Hoboken: John Wiley & Sons.

Tosteivn, G. (2019). Content Matters: the materiality of cultural transmission and the intersection of Paleolithic archaeology with cultural evolutionary theory. In, A.C. Love and W.C. Wimsatt (eds.), *Beyond the Meme: Articulating Dynamic Structures in Evolution*. Minnesota Studies in Philosophy of Science, pp. 311-364. Minneapolis: University of Minnesota Press.

Toulmin, S. E. (1958), The Uses of Argument, Cambridge: Cambridge University Press.

Toth, N., Schick, K. D., Savage-Rumbaugh, E. S., Sevcik, R. A., & Rumbaugh, D. M. (1993). Pan the tool-maker: investigations into the stone tool-making and tool-using capabilities of a bonobo (Pan paniscus). Journal of Archaeological Science, 20(1), 81-91.

Toth, N., & Schick, K. (2018). An overview of the cognitive implications of the Oldowan Industrial Complex. Azania: Archaeological Research in Africa, 53(1), 3-39.

Toth, N., & Schick, K. (2023). 95 Insights into the Cognitive Abilities of Oldowan and Acheulean Hominins: Experimental Approaches. In Oxford Handbook of Cognitive Archaeology (pp. 95-130). Oxford University Press.

Trinkaus, E., & Jelínek, J. (1997). Human remains from the Moravian Gravettian: the Dolní Věstonice 3 postcrania. *Journal of Human Evolution*, *33*(1), 33-82.

Trinkaus, E., & Buzhilova, A. P. (2018). Diversity and differential disposal of the dead at Sunghir. *Antiquity*, *92*(361), 7-21.

Turner, D. (2007). *Making prehistory: Historical science and the scientific realism debate*. Cambridge: Cambridge University Press.

Turner, D. (2011). *Paleontology: a philosophical introduction*. Cambridge University Press.

Turner, D. D. (2019). *Paleoaesthetics and the Practice of Paleontology*. Cambridge University Press.

Tucker, A. (2011). Historical science, over- and underdetermined: A study of Darwin's inference of origins. *British Journal for the Philosophy of Science*, 62(4), 805–829.

Turk, M., Turk, I., & Otte, M. (2020). The Neanderthal musical instrument from Divje babe I cave (Slovenia): a critical review of the discussion. *Applied Sciences*, *10*(4), 1226.

Tryon, C. & Bailey, S. (2013) Testing Models of Modern Human Origins with Archaeology and Anatomy. Nature Education Knowledge 4(3):4

Val, A. (2016). Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa?. *Journal of Human Evolution*, *96*, 145-148.

Vandermeersch, B., Bar-Yosef, O., 2019. The paleolithic burials at Qafzeh cave, Israel. PALEO, Revue d'arche´ ologie pre´ historique 30 (1), 256–275, http://dx.doi.org/10.4000/paleo.4848.

Vermeersch, P. M. (2002). Two Upper Palaeolithic Burials at Nazlet Khater. In: Vermeersch PM (Ed.) 2002. Palaeolithic Quarrying Sites in Upper and Middle Egypt. Leuven University Press: 273-282.

Villa, P., & Roebroeks, W. (2014). Neandertal demise: an archaeological analysis of the modern human superiority complex. *PLoS one*, *9*(4), e96424.

Villmoare, B., & Kimbel, W. (2024). On the scientific credibility of paleoanthropology. *Evolutionary Anthropology: Issues, News, and Reviews*, *33*(4), e22037.

Vrba, E. S. (2015). Role of environmental stimuli in hominid origins. In Henke, W. & Tattersall, I. (Eds.): *Handbook of paleoanthropology* (pp. 1837-1886). Springer, Berlin, Heidelberg.

Wagner, G. P. (Ed.). (2001). *The character concept in evolutionary biology*. Academic Press.

Wallach, E. (2019). Inference from absence: the case of archaeology. *Palgrave Communications*, *5*(1), 1-10.

Watkins, A. (2024). When Are Fossils Data?. Philosophy, Theory, and Practice in Biology, 16(2).

Westaway, K. E., Louys, J., Awe, R. D., Morwood, M. J., Price, G. J., Zhao, J. X., ... & Sulistyanto, B. (2017). An early modern human presence in Sumatra 73,000–63,000 years ago. Nature, 548(7667), 322-325.

Weston, E. M., & Lister, A. M. (2009). Insular dwarfism in hippos and a model for brain size reduction in Homo floresiensis. *Nature*, *459*(7243), 85-88. https://doi.org/10.1038/nature07922

White, R., Bosinski, G., Bourrillon, R., Clottes, J., Conkey, M. W., Rodriguez, S. C., ... & Willis, M. D. (2020). Still no archaeological evidence that Neanderthals created Iberian cave art. Journal of Human Evolution 144:102640. https://doi.org/10.1016/j.jhevol.2019.102640

White, T. D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C. O., Suwa, G., & WoldeGabriel, G. (2009). Ardipithecus ramidus and the paleobiology of early hominids. *science*, *326*(5949), 64-86.

White, T. D., Lovejoy, C. O., Asfaw, B., Carlson, J. P., & Suwa, G. (2015). Neither chimpanzee nor human, Ardipithecus reveals the surprising ancestry of both. *Proceedings of the National Academy of Sciences*, *112*(16), 4877-4884.

Wiersma, J. P., Roberts, E. M., & Dirks, P. H. (2020). Formation of mud clast breccias and the process of sedimentary autobrecciation in the hominin-bearing (Homo naledi) Rising Star Cave system, South Africa. *Sedimentology*, *67*(2), 897-919.

Wilson, E.O. (1975). *Sociobiology*, Cambridge, MA: Belknap/Harvard University Press

Wood, B., & Smith, R. J. (2022). Towards a more realistic interpretation of the human fossil record. *Quaternary Science Reviews*, *295*, 107722.

Wylie, A. (1985). The reaction against analogy. In *Advances in archaeological method and theory* (pp. 63-111). Academic Press.

Wynn, T. (2022). 'Redescribing the Oldowan', in Karenleigh A. Overmann, and Frederick L. Coolidge (eds), Oxford Handbook of Cognitive Archaeology, Oxford Library of Psychology (2024; online edn, Oxford Academic, 19 May 2022), https://doi.org/10.1093/oxfordhb/9780192895950.013.3

Wynn, T., & Coolidge, F. L. (2022). An introduction to evolutionary cognitive archaeology. Routledge.

Wynn, T., Hernandez-Aguilar, R. A., Marchant, L. F., & McGrew, W. C. (2011). “An ape's view of the Oldowan” revisited. Evolutionary Anthropology: Issues, News, and Reviews, 20(5), 181-197.

Wynn, T., & McGrew, W. C. (1989). An Ape's View of the Oldowan. Man, 24(3), 383–398. https://doi.org/10.2307/2802697

Wynn, T., Overmann, K. A., & Coolidge, F. L. (2016). The false dichotomy: a refutation of the Neandertal indistinguishability claim. Journal of Anthropological Sciences, 94, 1-22.

Zeller, E., Timmermann, A., Yun, K. S., Raia, P., Stein, K., & Ruan, J. (2023). Human adaptation to diverse biomes over the past 3 million years. *Science*, *380*(6645), 604-608.

Zilhão, J., Angelucci, D. E., Badal-García, E., d'Errico, F., Daniel, F., Dayet, L., ... & Zapata, J. (2010). Symbolic use of marine shells and mineral pigments by Iberian Neandertals. Proceedings of the National Academy of Sciences, 107(3), 1023-1028.

Zilhão, J. (2019). Tar adhesives, Neandertals, and the tyranny of the discontinuous mind. Proceedings of the National Academy of Sciences, 116(44), 21966-21968.