Citation for published item:

Pettitt, P. B. and Zilhao, J. (2015) 'Problematizing Bayesian approaches to prehistoric chronologies.', World archaeology., 47 (4). pp. 525-542.

Further information on publisher's website:

https://doi.org/10.1080/00438243.2015.1070082

Problematising Bayesian approaches to prehistoric chronologies

Paul Pettitt and João Zilhão

ABSTRACT

In recent years Bayesian exploration of radiocarbon datasets has been employed widely in prehistoric archaeology. Pertinent especially to major biogeographic and behavioural changes such as human dispersals and extinctions, the spread of agriculture and culture change, the method can offer a powerful means to improve considerably the precision of prehistorians' investigation of some of the most major questions in human prehistory. As such its potential is profound – it has even been regarded as the third radiocarbon revolution – but its appropriateness is dependent on the assumptions that must be made of the samples selected for dating. How sound are these assumptions, and therefore how reliable are Bayesian analyses? Here, we introduce some aspects and assumptions that underline Bayesian modelling of radiocarbon measurements, and we problematise their application in Palaeolithic archaeology. We conclude that many existing models are faulty, and suggest some criteria for quality control in this field.

INTRODUCTION

The last two decades have witnessed a major growth in scientific techniques available to archaeology. As machines and methods have developed, some techniques have ceased to be the preserve of trained specialists, and have become available to archaeologists at large. Archaeologists – far from being suspicious of 'science' as was often expressed in the 'processual versus postprocessual' debates that pervaded the 1970s-2000s – have, if anything, embraced it too enthusiastically or uncritically today (Killick 2015, 243). Amidst this floruit of archaeological science, radiocarbon dating has risen to dominate the chronometry of archaeological and palaeoenvironmental studies of the last 50,000 years (Wood 2015), and use of the Bayes theorem has become routine in the construction of radiocarbon-based chronological models in prehistoric and protohistoric archaeology. This 'explosion' of Bayesian modelling (Woods 2015, 67) has been referred to as the 'third radiocarbon revolution' (Bayliss 2009), and can be applied to other chronometric methods (Millard 2003). Although it developed out of the close collaboration between the archaeological ‘user community’ and statisticians, given the ready availability of Bayesian computer programs such as OxCal (e.g. Bronk Ramsey 2009a) it is now very easy for archaeologists to deploy Bayesian models themselves, without an understanding of the philosophical and mathematical assumptions on which they are based. As Hamilton et al. (this volume) note, however, building chronologies using sophisticated Bayesian modelling is a specialism that requires statistical training, and direct applications by non-specialists may run the risk that models become ‘black boxes’, used without any understanding of their mechanisms (Buck and Meson this volume).

This volume arose out of our concerns that a number of potentially flawed Bayesian models have been published in our own field – Middle and Upper Palaeolithic archaeology – the results of which profoundly affect our understanding of major issues such as the timing and nature of the extinction of the Neanderthals and the initial dispersals of anatomically modern humans across the Old World. If the Bayesian models are in any way incorrect, so are our reconstructions of these major palaeobiological processes. We wondered if our reservations were shared by archaeologists working in other periods, and we therefore sought to encourage debate about how models have been constructed and applied in prehistory over the last one and a half decades. Debate in any scientific field takes time and effort, and for our own field alone we were painfully aware of the voluminous literature generated over this time by chronological models that have been published for a small number of important sites. Such quality control issues are of course ubiquitous in science, although the archaeological community is perhaps in greater danger of uncritically accepting models and results that may be majorly flawed. As Killick (ibid., 246) has noted, we must try to filter out such flawed studies which have arisen from time to time as part of the ‘gorging’ of a ‘loaded buffet table of easy-to-use analytical techniques before archaeologists’ wherein ‘overindulgence has lead to indigestion, and rapid growth to growing pains’.

DEBATE ABOUT BAYESIAN MODELS

In archaeology, Bayesian models have been applied to site chronologies, human dispersals, artefact typological change and processes of cultural change, as well as climate-environment-demographic correlations (e.g. Riede and Edinborough 2012; Crema et al. 2014; see also Bayliss; Capuzzo and Barceló; Dye; Hamilton et al.; Teyssandier et al. and Banks, this volume for useful summaries). Beyond archaeological chronologies, Bayesian modelling has been employed extensively in biology, medicine, law, industry and computing. It has even been used to establish the probability of the existence of God (Unwin 2003). The archaeological floruit is not surprising given that several programmes for Bayesian modelling are freely available online (see Woods 2015, 67). When used correctly these are powerful tools which may improve chronometric precision significantly, eliminate uncertainties and outliers (Bronk Ramsey 2009b), and provide independent verification of chronologies established by other means (see, for example, Hamilton et al. this volume for the British Iron Age). But how do we know when they are being used correctly (or in the words of Buck and Meson, this volume, who is being a 'good Bayesian')?

Bayesian models may be appropriate for relatively straightforward archaeological questions where pertinent priors are available. An example of this is the comparison of the timing and rate of replacement of inhumation by 'Urnfield' cremation cemeteries in Switzerland and Catalonia during the second millennium BC (Capuzzo and Barceló, this volume). This analysis reveals a different archaeological (and one therefore assumes, social) signature in each region. Such studies show how simple models can inform on complex processes of archaeological change.

In our own field, debate over Bayesian modelling has been intense, and relates to two issues; whether the assumptions that underlie Bayesian models are justifiable; and whether 'prior' information essential to their construction (in this case, samples that have been dated) is relevant and reliable. The contributions to this volume, which range in time from the Middle Palaeolithic to the Iron Age, address the former (Buck and Meson; Weninger et al.) and latter (Banks; Bayliss; Discamps et al., Hamilton et al., Capuzzo and Barceló). In terms of sample selection Bayliss (this volume) considers most projects as 'generally flawed', which is a worrying state of affairs.

Statistics

As Buck and Meson (this volume) note, describing oneself as a 'Bayesian' is actually saying some very specific things about one's approach to problems, methodologies and interpretations, and simply using Bayesian software does not necessarily make research Bayesian, or at least good Bayesian work. Buck and Meson approach this issue from the wider context of the philosophy of science. Their sobering critique – particularly about the selection of appropriate priors and about the complex task of combining probability distributions for all but the simplest models - makes it imperative that we continue to foster meaningful collaborations between archaeologists and statisticians.

The need for such collaboration is ably shown by Weninger et al. Their deconstruction of the process of calibration reveals what visually-pleasing age plots actually conceal; lines between uncalibrated and calibrated age ranges on curves need not reflect identical things just because they look identical. This raises important questions about the assumptions which underlie the process of calibration, and thus of Bayesian modelling. While the most commonly-used calibration programmes produce calendar age results that are very similar, with large datasets even the smallest fault in assumptions can be magnified greatly.

Priors

The main responsibility of archaeological users of Bayesian models is to ensure the reliability of the *prior* information upon which models are based, which equates to establishing beyond reasonable doubt the stratigraphic integrity of the samples from which the modelled dates derive, and their pertinence to the archaeological question at hand. Essentially, the technique allows users to combine measured data from a sample (in this case an age range of a dated sample, expressed as a probability distribution) with knowledge about the context of the sample before (*prior* to) its age measurement. If there is any doubt about this prior information - if the priors are 'corrupt' in the terminology of Weninger et al. (2010, 982) - the models can introduce significant inaccuracy into the resulting age model or *posterior* probability distributions (Steier and Rom 2000, 183). Thus 'any problems with the samples, their context, their association with each other…will have implications for the accuracy of our chronologies' (Bronk Ramsey 2009a, 358). It follows that honest and careful scrutiny of the constituent parts of all Bayesian models should be routine: 'priors can have a significant effect on the posterior distributions; thus, they must be used judiciously' (Higham et al. 2010, 20235). The implications of priors should be clear; without critical assessment of their stratigraphic/contextual reliability, results can be misleading or simply plain wrong.

As Dawkins (2007, 133) has noted, priors are usually subjectively selected, and as such can be judged on the basis of the *GIGO* principle; Garbage in, Garbage out. Select the wrong priors and how would one know if the results are meaningful? Because of the assumptions and subjectivity that often underlie the selection of priors the method is 'avowedly subjective' (Buck and Meson, this volume). Objectively assessing the reliability of priors is not an easy task, however; it has been described as 'the hardest part of Bayesian analysis' (Bronk Ramsey 2000, 201). Yet 'since in many cases assumptions/guesses are needed to construct these probability functions, it is important to assure [sic] that these assumptions/guesses do not bias the results' (Steier et al. 2001, 379). The problem with priors is discussed in all contributions to this volume. To Buck and Meson the first requirement of a conscientious Bayesian is to have a good intuitive feel for the model and its parameters, achieved by meaningful engagement with the model and its writer as well as with the priors. These are not empty terms: *engagement* means critically evaluating the set of priors and their potential effects on the results, and only by so doing can one claim to be aware of the *parameters*. The key is to evaluate the reliability of results by varying our likelihoods and priors within a reasonable range and to compare the results. Such exploration is noticeably lacking in almost all Bayesian models in archaeology, however. As Bronk Ramsey (2000, 201) notes, 'there is no one correct prior for a given situation. A prior is merely a model that can be applied to the data to help in its interpretation. Ideally, several different models with different priors should be tried…this approach may not appeal to some people who wish to take their results, process them statistically and come out with the "right" answer' (a view also expressed, for example by Steier et al. 2001, 379). Discamps et al. (this volume) present several potent examples where extreme caution needs to be applied to the selection of pertinent dated samples in Bayesian datasets for Middle and Upper Palaeolithic archaeology, and show how conclusions based on questionable samples can be very wide of the mark. As they note, evaluating the taphonomy and relevance of archaeological samples may be more pressing a problem than evaluating the results of the Bayesian models based upon them.

A major problem, in our opinion, is that priors are often selected uncritically. This is perhaps not surprising; for many sites the range of available samples for dating is limited; if the context of available samples was scrutinised critically few would be found appropriate, and sites, therefore, would not be susceptible to Bayesian analysis. Thus, dubious datasets form the priors for ambitious questions. Following modelling, posteriors (results) are then often accepted uncritically; prior and posterior data are accepted at face value. An example of this can be seen in the broad acceptance of artefact-based chronologies for the British Iron Age, which is unwarranted, as the assumed age of contexts from which the artefacts derive has not been adequately demonstrated through dating of multiple sites, as Hamilton et al. (this volume) note.

Dye (this volume) uses Bayesian modelling to explore the pattern of the initial settlement of the Pacific Islands, noting with the example of Hawai'i the problem of a *disparity* between the timing of the event one is trying to establish (e.g. a first appearance of humans in a given region) and the apparent timing of the event in terms of the evidence one has available. His example shows how assumptions about the scale of such disparities form the main source of disagreement over chronologies. As it is impossible *a priori* to establish whether a real ('target') date pre- or post-dates an apparent one, Bayesian models can form a useful way of dealing with such disparity, although an important lesson from the history of research into the settlement of Hawai'i is that it is impossible to ascertain how strict or how lax rejection protocols should be, which has taken estimates from too young to too old before arriving at the modern consensus.

RECENT DEBATES IN PALAEOLITHIC ARCHAEOLOGY

Radiocarbon dating of archaeological assemblages as well as climatic and environmental change plays a critical role in late Middle and Upper Palaeolithic archaeology (e.g. Conard & Bolus, 2003; d'Errico & Sánchez-Goñi, 2003; d'Errico et al., 2012 *contra* Benazzi et al., 2011; Fedele et al., 2008; Jöris & Street, 2008; McPherron et al. 2012; Mellars, 2004; Roebroeks, 2008; Talamo et al. 2012a, b; Tzedakis et al., 2007; Walker et al., 2008; Zilhão, 2006, 2013; Zilhão & d'Errico, 1999; Zilhão et al., 2010, 2011a, b). Given that many ambitious uses of large scale chronometric datasets can be described as 'somewhat cavalier' (Roebroeks 2008, 918) it is no surprise that a sizable literature has accumulated surrounding the controversies over the dating of sites with deep stratigraphies that have so far been crucial to our understanding of the Middle to Upper Palaeolithic transition, and most of this debate has centred around a relatively few of these. Accurate and precise chronological control is critical to research in the period, which relates to the biogeography of Neanderthal extinction and modern human dispersal, and to the elucidation of their climatic and environmental contexts. These complex events, played out at a Eurasian scale over several millennia of Marine Isotope Stage (MIS) 3 'mark crucial developments in human biological and cultural history and are among the most debated issues in palaeoanthropology and archaeology' (Jöris and Street 2008, 782), yet as Banks (this volume) notes, much needs to be done in order to extend studies of climatic correlation away from single site sequences to wider datasets.

The perceived stratigraphic uncertainties and taphonomic issues that bear upon the potential relevance of dated samples to the questions of concern (often about the last appearance of Neanderthals and/or the first appearance of anatomically modern humans at a given site or in a given region) are conspicuous problems. The sites central to these debates were often excavated some time ago, without the benefit of current excavation and recording methods, and without modern research questions in mind. In some cases, dates obtained on large numbers of humanly-modified artefacts from deep stratigraphies can appear to be relatively unproblematic (e.g. Riparo Mochi, Italy: Douka et al. 2012; Les Cottés, France: Talamo et al. 2012a; Abri Pataud, France: Higham et al. 2011d), although modern pretreatment and dating of samples can still reveal many problems with context and contamination (a good example being the case of the Grotta di Fumane, Italy:

Higham et al. 2009). Sites where either young radiocarbon dates have been interpreted as reflecting late Neanderthal survival in a given region or at which the Aurignacian and Châtelperronian have been seen (erroneously) to be interstratified, played the major role in the suggestion that Neanderthals and modern humans were in some regions chronometrically contemporary (not that this would necessarily equate with real contemporaneity however). Examples of the former include Gorham's Cave, Gibraltar (Finlayson et al., 2006, 2008 *contra* Zilhão & Pettitt, 2006) the Grotte du Renne at Arcy-sur-Cure (Higham et al., 2010, 2011a, b, *contra* Caron et al. 2011, Zilhão et al. 2011a, b) and of the latter the Grotte des Fées at Châtelperron (Mellars, 1999; Gravina et al. 2005 and Mellars et al. 2007 *contra* Zilhão et al. 2006, 2008a, b).

Overall, given the problems of accuracy and precision often relating to sample context or pretreatment/measurement, as well as a relatively imprecise calibration curve, the chronology of MIS3 can be characterised as poor (Pettitt et al. 2003; Zilhão and d'Errico 2003a), and it is no surprise that understanding the transition has been seen as 'one of the most major intellectual challenges in archaeology' (Talamo et al. 2012b). The misuse or misinterpretation of dates can therefore be highly misleading, and in such contexts, single measurements or modelled sequences based on a small number of dates from taphonomically complex sites can change the field if they are accepted uncritically. To archaeologists lacking training in statistics, Bayesian models are difficult to understand; but while they may not be able to judge the statistical rationale underlying models, they are capable of assessing them on the basis of the priors that they are built upon. For Palaeolithic sites which accumulate in natural sediments, archaeological 'levels' are often thick palimpsests, and their formation was usually complex; cultural 'layers' are often arbitrary creations of the excavator and artefacts may be mobile between these. In such cases, what do 'outliers' actually mean? In Bayesian models these are usually assumed to be stratigraphically residual or intrusive items and are thus omitted, but they could equally reflect significantly long periods of sedimentary formation with episodes of deflation and therefore should not be 'written off' without further elucidation of site taphonomy. For these reasons the selection of suitable priors is often highly problematic, and we have ourselves been critical of several models applied to the Middle to Upper Palaeolithic transition.

Demonstrating that priors are not corrupt is therefore a particularly pronounced problem in Palaeolithic archaeology, and if this cannot be done, then the garbage going in becomes the garbage coming out. As Discamps et al. (this volume) reveal in their examination of the transition, poor sampling can seriously distort age estimations. One can only eliminate such problems with large series of dates (in which 'outliers' – whatever they mean - should be apparent); 'Experience has shown that to investigate the chronology of sites dating to the Palaeolithic properly, one requires a large series of [dated] samples' (Higham et al. 2010, 20235). Then, it is usually assumed that outliers indicate stratigraphic mobility; 'clearly, it is important to demonstrate stratigraphic integrity in any archaeological context. One useful way to do this is with a series of well-selected radiocarbon dates from throughout a succession of archaeological strata. Variation in the results outside that expected statistically might be held to herald problems within the sequence, whereas the reverse would improve confidence in its integrity' (Higham et al. 2010, 20235). This should be obvious, but does not in all cases deter the use of Bayesian age models to calculate phase boundaries on the basis of very small numbers of dated samples (e.g. Pech de l'Azé IV – McPherron et al. 2012). These may be correct, but if larger numbers of samples are measured and included in the models would the results change? Another source of error is when single layers are represented only by one radiocarbon measurement. How do we know if this is accurate? Sadly, it is also apparent that Bayesian models only very rarely explore the priors in this way.

In Palaeolithic archaeology, therefore, Bayesian models have too frequently been constructed on the basis of questionable sets of priors, and have not been used to explore or evaluate the reliability of possible outcomes. Instead, major conclusions based on such models are published in major journals and can be accepted by the scholarly community at face value. The resulting debate has been intense, and we wonder whether any other area of research into the Palaeolithic has received as much argument. Not that interpretation is free of error either. As Verpoorte (2005, 274) has noted, from a range of dates produced by Bayesian models, it is the oldest that receive most attention, and are usually interpreted as the most reliable. It should be obvious how such 'dragging' (in the terminology of Discamps et al. this volume) potentially introduces serious errors into research aimed at establishing the 'earliest' events such as the dispersal of anatomically modern humans across Europe (e.g. Higham et al. 2011c). Proponents of these models need to be open about the inbuilt biases in the statistical technique.

Not that existing debate has been particularly helpful; critiques are often met with hostile ripostes, some of which amount to little more than arguments from authority which are not backed up by reasoned argument. As one of us has noted before, these should have no place in scientific debate (Zilhão et al. 2011). One of the major proponents of modelling in this field, for example, blames critics for their lack of understanding; 'it becomes clear...that Caron et al. do not understand the basis of the [Bayesian] method' (Higham et al. 2011/12, 6); and Higham and Stringer (2012) suggested that the conclusions of White and Pettitt (2012) – who criticised their 'dating' of a human maxilla from Kent's Cavern, UK - 'expose a breathtaking ignorance of the developments in scientific approaches to the past' (quoted at http://news.sciencemag.org/sciencenow/2012/08/the-mysterious-affair-at-kents.html, last accessed 16/04/2015). Such responses are unproductive at best, and certainly do not address the critiques. We believe that the onus is upon the proponents of Bayesian modelling to explain their methods explicitly, justify their priors and posteriors, explore the parameters of their models, and enter into fair and reasoned debate with specialists whose views are at odds with theirs.

A balanced and open debate about the use of such a potentially powerful technique is therefore long overdue. In order to try to stimulate this process in our own field, we isolate the most pressing desiderata of Bayesian models, using examples to show how problems occur and how debate has progressed.

Research question: is the research question amenable to Bayesian modelling, and does it make sense to begin with?

An example of Bayesian analysis as an end in itself is a recent examination of the chronology of the Portuguese Gravettian (Bicho et al. 2014), which asks no archaeological question other than 'what dates are these?', and as a result gains no new knowledge. The specific aim is to establish 'a tight and better estimate with a high statistical significance for the chronometric sequence and *the different boundaries for the beginning and ending of each Gravettian phase*' (ibid., 505, our emphasis). In addition to establishing the beginning and end of the Gravettian, their purpose is to establish a 'boundary' (i.e. time of transition) for the 'interface' (whatever that means) between an 'Early Gravettian' and 'Late Gravettian'. It is apparent, however, that these entities are defined not on technological or typological grounds but on chronological ones: 'the earliest phase is not well represented…there are only two sites associated with this chronology…a third site has a long chronology including various layers dating to this time, but no unequivocal lithic artefacts can yet be attributed to the Early Gravettian…the later phase has a wide diversity…dependent on the region and possibly chronology, although no true data exists to confirm this hypothesis (ibid., 500). Thus a 'chronological interface' is defined on the basis of a scanty chronology, and is therefore a mathematical artefact for which one cannot – as yet at any rate - demonstrate any archaeological significance whatsoever. No Bayesian analysis - actually no modelling of any kind - is necessary to conclude that, for a number of dates falling in a given time interval, some will define an earlier range and the others will define a later range.

Sample selection: dated samples must reflect human activity

Including dates on samples from both humanly- and carnivore-modified bones will inevitably introduce error into models, as in the case for establishing Neanderthal presence at Les Cottés cave, France (discussed by Discamps et al. this volume). As they show, when including only dates on humanly-modified materials in the model, the age for the Mousterian (Neanderthal presence) at the site becomes older, removing the possibility of 'dragging' and therefore any suggestion of it being contemporary with anatomically modern humans.

More dubiously still, Higham et al. (2011c) dated a number of unmodified faunal remains derived from poor excavations of the 1920s in Kent's Cavern (Torquay, UK) in order to 'date' a human maxilla fragment tentatively identified as modern human which could not be dated directly. None of the dates therefore reflect human activity, but if in reliable stratigraphic order bracket the maxilla with minimum and maximum ages on samples recovered from above and below it. If their results were at all reliable, they would indicate a significantly early arrival of anatomically modern humans in northwestern Europe, at a time when Neanderthals were probably still extant in Belgium. The cave's deposits derived from high energy sedimentation, were subsequently disturbed, poorly excavated, and sparsely recorded (White and Pettitt 2012). We note that Kent's Cavern is omitted from discussion of potential contemporaneity between Neanderthals and modern humans in Higham et al. 2014. The lesson is that unless one has humanly-modified materials available for dating, some early excavations should be left well alone.

Site formation: samples must derive from reliable stratigraphic sequences and post-depositional disturbance should be eliminated

Collections from old excavations, for which methods of recovery and recording may be poor relative to those of today, must be treated with extreme caution. Pirson et al. (2012) undertook a major dating project of several Belgian caves pertinent to the Middle to Upper Palaeolithic transition, ruling out several for such historical reasons. On the basis of a major AMS radiocarbon dating of samples from the Middle and Upper Palaeolithic levels of the Grotte du Renne, Arcy-sur-Cure (Yonne, France) Higham et al. (2011a) concluded that stratigraphic mixing had occurred, and thus that one could not be confident that the apparent association between Neanderthal remains and the site's Châtelperronian archaeology was reliable. Some took this to imply that the apparent association of symbolic artefacts and the Neanderthal remains was also accidental (e.g. Mellars 2010). Subsequently, Caron et al. (2011), Zilhão et al. (2011a) and Hublin et al. (2012) demonstrated that such mixing was minimal at worst. Higham et al. (2011b) defended their laboratory techniques, maintaining that taphonomy best explained their outliers, although failed to account for strong arguments as to a generally low level of arfefact mobility (Zilhão et al. 2011a).

Benazzi et al. (2011) relied on eight dated shells to develop a Bayesian model for the Uluzzian stratigraphy of the Grotta del Cavallo in southern Italy, in order to date two teeth which they identify as modern human (that this identification is debatable will not be discussed here). Although some of the shells have been modified into personal ornaments their stratigraphic relationship with the teeth is unclear. One is acknowledged to represent 'an old shell, either collected on purpose, or accidentally brought to the site' (ibid., supplementary information). The others come from deposits affected by major post-depositional disturbance, acknowledged by the excavator and made evident by the ca.19 ka date obtained for a sample considered to be in situ in the Uluzzian. Given this, how can one reliably establish a chronological relationship between the ornamental shell and the human remains they are supposed to date by association (Zilhão 2013)? The interpretation that arose from this research redefined the entire Uluzzian as a product of modern humans, not Neanderthals, and by so doing back dated the arrival of *Homo sapiens* in Italy (Douka et al. 2014; Moroni et al. 2012).

Site excavation: levels from which samples derive should be meaningful, not arbitrary

Conard and Bolus (2003) suggested a relatively early age for the appearance of the Aurignacian (and thus anatomically modern humans) in the Swabian Jura on the basis of AMS radiocarbon measurements from Das Geissenklösterle. Despite a strong critique based on stratigraphic observations (Teyssandier 2003; Zilhão and d'Errico 2003b; Verpoorte 2005), Higham et al. (2012) found support for this early age on the basis of Bayesian modelling of 23 new AMS radiocarbon measurements. The problem derives from the stratigraphic significance of the levels into which the excavator divided the relevant sediments, in addition to demonstrable post-depositional disturbance, and the use of samples that do not clearly relate to human activity, which indicate that the interface between an Upper Aurignacian level IIb and lower Aurignacian level III is not a discrete boundary between distinct taxonomic units (Banks et al. 2013; Zilhão 2013; Discamps et al., this volume). The result is a logically invalid set of priors for the Bayesian modelling.

Interpretation: the meaning of 'boundaries' in models should be understood

Despite the stratigraphic problems with Das Geissenklösterle, the results (Higham et al. 2012, Table 3) were interpreted as indicating the presence of an Early Aurignacian (and thus anatomically modern humans) at the site around 42,500 cal BP, this estimate deriving from the modelled transition between horizon AHIV (the uppermost Middle Palaeolithic) and the overlying Aurignacian in horizon AHIII. To an extent this estimate could be a factor of 'dragging', i.e. one in which the parameters of the model itself result – in this case – in a relative old age. 'Boundaries' in Bayesian models are far more than convenient 'breaks' between dated horizons: stratigraphic assumptions are made when they are inserted into models, and although these may be implicitly logical, they can affect interpretation. It is important to remember that boundaries provide *maximum* ages for the overlying archaeology, as is the case for the 'AHIII Start Boundary' at Geissenklösterle. Thus, based solely on humanly-modified materials in AHIII and AHIV and unmodified bones in the archaeological sterile horizon between the two, Higham et al.'s age range of 43,060 – 41,480 cal BP (2σ) for the boundary indicates that the *maximum* age for the site's Aurignacian is no more than 43,060 and, taking the modelled age range into account, this could be younger than 41,480 cal BP. The age ranges of the two measured samples from the sterile layer overlap between 42,110 and 44,330 cal BP (2σ) and the modelled age range for this 'level' is 41,860-43,10cal BP (2σ). Again, it is important to remember that this age range does not pertain to the *presence* of the (overlying) Aurignacian at the site, but provides a maximum age for it. Despite this, Higham et al. conclude that the Aurignacian was present at the site around 42,500 cal BP. This conclusion is not necessarily inconsistent with the maximum age range, but as this could be as young as 41,480 cal BP (and the Aurignacian therefore younger) it is a good example of how Bayesian modelling can drag the age backwards.

Interpretation: archaeological proxies for Neanderthals and modern humans must be reliable.

Camps and Higham (2012) attempted to date the timing of the Middle to Upper Palaeolithic transition at the Abric Romaní (Catalunya, Spain). Central to their argument is the conclusion that the earliest Upper Palaeolithic level – Layer A – is a mix of Aurignacian I and later archaeology. Despite this, they assume that perforated shells (of two distinct genera) from Layer A 'are approximately the same age', and thus assume a minimum age for these of ~36 ka (^{14}C) BP (the age of the oldest measurement), rejecting two measurements that are considerably younger. Given that

U-series measurements on tufas that formed below Layer A and charcoal measurements from Layer A have relatively large errors, this assumption is critical to the subsequent Bayesian model that Camps and Higham construct. Despite 'several outliers of significance' in this (ibid., 98) they conclude that modern humans had appeared in the region some 41.2-42.3 ka (cal) BP (2σ), 'some 1500 years prior to the H4 event' from which time many date the arrival of modern humans. Their identification of priors, however, has been shown to be 'seriously flawed' due to a number of inaccuracies in their classification of lithic artefacts from both late Middle and early Upper Palaeolithic layers (Vaquero and Carbonell 2012, 713). In addition to redefining the site's late Middle Palaeolithic technologically, Camps and Higham apparently identified a number of Châtelperron points, which they have never illustrated but which would make the site a richer Châtelperron collection than several well-known examples. Most of their assumptions and identifications are incorrect as Vaquero and Carbonell make clear, and thus Camps and Higham's interpretation of the resulting Bayesian model lacks archaeological significance.

Interpretation: age ranges of dated samples should be interpreted in the light of stratigraphic priors

Both Discamps et al. and Banks (this volume) provide strong critiques of how Bayesian modelling has been applied to the Middle to Upper Palaeolithic transition, and conclude that existing models are often highly questionable. Where others have interpreted apparent chronological overlap between the Châtelperronian and Mousterian as reflecting contemporaneity between the two, if one accounts for the simple stratigraphic fact that on all sites where the two co-exist the former always follows the latter, it becomes clear that apparent overlap between the two in chronological models where this prior information has not been given sufficient attention, is simply an inevitable consequence of the uncertainty (i.e. error ranges) associated with radiocarbon measurements, potentially exacerbated greatly by long periods of sedimentary formation separated by episodes of deflation. Thus, even though the Châtelperronian *follows* the Mousterian, their age ranges will overlap even when the site's stratigraphic integrity and the pretreatment and measurement of samples from it are ideal. The dating of stratigraphic units in this time period is resolvable to a millennium at best (often worse), and given the problems of stratigraphic mobility of samples, the presence of residual and/or invasive priors should be considered to be the rule, even if no obvious post-depositional disturbance seems to have affected particular sites. Thus Higham et al.'s (2014) interpretation - that their model's indication of chronological overlap is proof of contemporaneity - needs not be the case at all, and is logically unsupportable.

Accuracy: despite recent progress, the accuracy of results close to the method’s limit of applicability cannot be taken for granted and is simply another prior in the model, one that needs to be explicitly considered and parametrized

For the Middle to Upper Palaeolithic transition, age ranges of radiocarbon measurements were often beyond the range of the INTCAL04 calibration curve, although may fall into the range of IntCal09 and its successor, IntCal13 (Reimer et al. 2009, 2013). A number of factors may account for age over-estimation or imprecision in the construction of calibration curves (see discussion in Reimer et al. 2013, 1870-3). Prior to IntCal13 and IntCal09, the IntCal Working Group ‘deemed the discrepancy among even the most robust datasets [in IntCal04] too large to make a reliable calibration curve beyond 26,000 cal BP’ (Reimer et al. 2009, 1114). The subsequent ‘NotCal’ curve contained major offsets, which were eliminated in the next (IntCal09) curve. This demonstrates that the accuracy of such curves can be improved over time, although this is an ongoing process and ‘indicate[s] that the shape of the present day calibration may still change’ (ibid., 1114). With this caveat in mind, ~~and~~ IntCal09 was issued as an interim curve, and although ~~with this caveat.~~ further ~~I~~improvement is noticeable ~~with~~ in IntCal13 ~~but~~ imprecision and inaccuracy still remain.

It is not our purpose here to critique the accuracy and precision of IntCal13, but to note that the potential inaccuracies of this or any calibration curve currently available may be exacerbated by the Bayesian modelling of age ranges produced using these curves. We worry whether Bayesian models will be biased towards the oldest possible results that age ranges suggest, and have discussed this above in respect to the modelling that was undertaken for Das Geissenklösterle, using data from Higham et al. (2012: actually based on IntCal09).

In addition to these potential effects of modelling, we must also note how variable the success of sample pretreatment has been in removing contaminating carbon. While it is not our remit here to discuss issues of decontamination (and, it must be noted that significant advances have been made in pretreatment methodology in recent years – e.g. Higham et al. 2006; Talamo and Richards 2011), it does remain a problem that relatively minor inaccuracies produced by insufficient decontamination could have major effects on calibration and subsequent age models (and we disagree with the contention of Higham et al. (2001b, 4) that differences of around 1100 cal years are ‘marginal’). Despite considerable advances in the development of the IntCal curves, the extremely low amounts of remaining ^{14}C beyond five half-lives means that ‘^{14}C dating of samples older than 30,000 years is still challenging and requires *outstanding* efforts in sample selection and in laboratory procedures' (Talamo et al. 2012b, 2466, our emphasis).

Decontamination of bone - which provides the main sample material for the transition - is especially problematic in this time range (Talamo and Richards 2011). Contamination has therefore formed a focus of debate about the reasons for discrepancies between the age models of Higham et al. (2010) and Hublin et al. (2012) for the Middle to Upper Palaeolithic transition at the Grotte du Renne at Arcy-sur-Cure. The potential effects of contamination were first raised by Caron et al. (2011, 5-7), who noted that 'the potential impact that even trace amounts of contaminants have on samples whose ages are close to the limit of applicability of radiocarbon is huge', and questioned several results on such grounds. It is known that faunal remains and bone tools from the site were extensively treated with preservatives including glues (Zilhão et al. 2011a, 3). Higham et al. (2011a, 2) expressed confidence in their assumption that this was not a problem, as although the 'bone industry…has visible evidence of consolidation by artificial conservation products, the same cannot be said of the faunal remains.' This argument from authority is expressed, despite two co-authors of Zilhão et al. (2011a) knowing the curatorial history of the collection; if Higham et al. were relying solely on visual inspection of these samples, it is in no way a reliable means of establishing whether or not a sample has been conserved. Although they took the precautionary measure of an additional solvent wash on these samples, a detailed debate followed over the veracity of their pretreatment and the potential effects of any unsuccessful decontamination on their results (see references above and particularly Higham et al. 2011a, b; Zilhão et al. 2011a). An obvious problem here is when samples appear to be/are known to be treated with conservatives, and/or results seem in some way odd, but which are nevertheless accepted 'with caution' (Higham et al. 2011a, 5), or authors 'err on the side of caution' stating that they are not 100% certain of results but include them anyway (ibid., 5). Why introduce potentially misleading data into publications? Doing so inevitably opens one to justifiable critique. Let us be clear what statements of this kind actually mean: 'this result could well be garbage; we don't know, but we are publishing it anyway. If it is garbage, however, so is the model we incorporate it in. We may well try to verify it later, but unless it is criticised, we probably won't.'

It should be clear that one simply cannot justify including any doubtful or debatable results in Bayesian models (see Buck and Meson, and Bayliss this volume). We suggest that in any cases of doubt, the onus must surely be upon the proponents of the dating project to avoid any potential contamination, or at the very least make it routine that they justify very strongly why conserved samples have been included in the modelling. Subsequent debate may clarify issues somewhat, but such quality control should not be the responsibility of others, and the best way to save people's time is to avoid any doubt completely by undertaking stringent quality control at the outset. For old excavations the lack of adequate recording or curatorial history may mean that it is impossible to vet samples in a satisfactory way for their stratigraphic and chemical integrity. If this is so, then we must avoid them, however disappointing this may be.

CONCLUSIONS: GARBAGE IN, GARBAGE OUT

The papers in this volume problematise a variety of assumptions critical to the use of Bayesian models in archaeology. Some provide examples of how 'good' models can be constructed; most raise more problems with how modelling has been used over the last decade or more. The main lesson, we think, is that quality control has often been conspicuously lacking from many published models. Perhaps this is no surprise, as a number of subfields of archaeological science have as yet inadequate quality control (Killick 2015, 243). But this is no excuse, nor is the fact that quality control may emerge through subsequent debate. Materials from old excavations may simply not meet anywhere near the stringent criteria for adequate construction of priors; in such cases they should be left alone. We need to identify the minimum conditions that should be met in any 'good' Bayesian modelling for archaeological chronology, and we hope this volume will begin a discussion in this light. We suggest that the following minimal conditions should apply:

- A research question that reflects archaeological concerns, rather than Bayesian modelling for its own sake
- Robust pretreatment methods for all samples included in models, and full delineation of pretreatment in publications; avoiding samples with any potential for contamination
- Modelling only with dated samples that are demonstrably pertinent to the research question at hand, such as human remains or other organic remains bearing obvious traces of human modification, not those with dubious stratigraphic associations
- Demonstrable stratigraphic integrity of the sediments from which samples have been selected for dating, meaningful subdivisions that reflect archaeological reality rather than arbitrary recording, lack of pronounced artefact mobility between such subdivisions, and incorporation of refitting studies, sediment micromorphology and other relevant methods in order to understand the nature of sedimentary units as clearly as possible
- Explicit justification of all priors based on the criteria above

- Presenting clearly explicable modelled sequences based on this accumulated quality control
- Exploration of potentially distinct (and potentially mutually exclusive) outcomes of Bayesian models and their implications for the research questions at hand, e.g. through consideration of the effects of 'dragging', and considering how different calibration curves affect the modelled results
- Open and fair engagement in subsequent scientific debate, without recourse to arguments from authority.

We hope that these suggestions will seem obvious components of good science and 'good citizenship' of the research community. Unless archaeologists engage critically with the priors upon which they base their Bayesian models, and the effects of calibration and subsequent modelling on these, they are in danger of producing results that are misleading at best, and which may potentially skew seriously our understanding of the major climatic, environmental and demographic processes we are interested in. Based on the principle of 'garbage in, garbage out' we must surely treat even the simplest of models as speculative, and therefore subject to scrutiny.

ACKNOWLEDGMENTS

We are grateful to the contributors to this volume who engaged critically with the theme. We are particularly thankful to our anonymous reviewers, who gave up their precious time to give close and constructive attention to the contributions. They know who they are! For our own contribution we are grateful to Manuel Vaquero and two anonymous referees who helped to improve the manuscript.

REFERENCES

Banks, W. E., d'Errico, F. and Zilhão, J. 2013. Revisiting the chronology of the Proto-Aurignacian and the Early Aurignacian in Europe: a reply to Higham et al.'s reply to Banks et al. (2013). *Journal of Human Evolution* 64, 39-55.

Bayliss, A. 2009. Rolling out revolution: using radiocarbon dating in archaeology. *Radiocarbon* 51(1), 123-47.

Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., Condemi, S., Ronchitelli, A., Harvati, K. & Weber, G.W. 2011. Early Dispersal of Modern Humans in Europe and Implications for Neanderthal Behaviour. *Nature*, 479: 525-28.

Bronk Ramsey, C. 2000. Comment on ‘the use of Bayesian statistics for ^{14}C analysis of chronologically ordered samples: a critical analysis’. *Radiocarbon* 42(2), 199-202.

Bronk Ramsey, C. 2009a. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1), 337-60.

Bronk Ramsey, C. 2009b. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51(3), 1023-45.

Camps, M. and Higham, T. 2012. Chronology of the Middle to Upper Palaeolithic transition at Abric Romaní, Catalunya. *Journal of Human Evolution* 62, 89-103.

Caron, F., d’Errico, F., Del Moral, P., Santos, F. & Zilhão, J. 2011. The Reality of Neanderthal Symbolic Behaviour at the Grotte du Renne, Arcy-sur-Cure, France. *PLoS ONE*, 6: e21545.

Conard, N. & Bolus, M. 2003. Radiocarbon Dating the Appearance of Modern Humans and Timing of Cultural innovations in Europe: new results and new challenges. *Journal of Human Evolution*, 44: 331-71.

Crema, E. R., Edinborough, K., Kerig, T. and Shennan, S. J. 2014. An approximate Bayesian computation approach for inferring patterns of cultural evolutionary change. *Journal of Archaeological Science* 50 (160-70).

Dawkins, R. 2007. *The God Delusion.* London: Black Swan.

Douka, K., Grimaldi, S., Boschian, G., del Lucchese, A. and Higham, T. F. G. 2012. A new chronostratigraphic framework for the Upper Palaeolithic of Riparo Mochi (Italy). *Journal of Human Evolution* 62, 286-99.

Douka, K., Higham, T. F. G., Wood, R., Boscato, P., Gambassini, P., Karkanas, P., Peresani, M. and Ronchitelli, A. M. 2014. On the chronology of the Uluzzian. *Journal of Human Evolution* 68, 1-13.

d’Errico, F. and Sánchez-Goñi, M.F. 2003. Neanderthal Extinction and the Millennial Scale Climatic Variability of OIS3. *Quaternary Science Reviews,* 22: 769-88.

d’Errico, F., Borgia, V. & Ronchitelli, A. 2012. Uluzzian Bone Technology and its Implications for the Origin of Behavioural Modernity. *Quaternary International* 259, 59-71.

Fedele, F.G., Giacco, B. & Hajdas, I. 2008. Timescales and Cultural Process at 40,000 BP in the Light of the Campanian Ignimbrite Eruption, Western Eurasia. *Journal of Human Evolution,* 55: 834-57.

Finlayson, C., Giles Pacheco, F., Rodríguez-Vidal, J., Fa, D.A., Guiterrez López, J.M., Santiago Pérez, A., Finlayson, G., Allue, E., Baena Preysler, J., Cáceres, I., Carrión, J. S., Fernaández- Jalvo, Y., Gleed-

Owen, C.P., Jimenez Espejo, F.J., López, P., López Sáez, J.A., Riquelme Cantal, J.A., Sánchez Marco, A., Giles Guzman, F., Brown, K., Fuentez, N., Valarino, C. A., Villalpando, A., Stringer, C.B., Martinez Ruiz, F. & Sakamoto, T. 2006. Late Survival of Neanderthals at the Southernmost Extreme of Europe. *Nature,* 443: 850-3.

Finlayson, C., Fa, D.A., Jimenez Espejo, F.J., Carrión, J.S., Finlayson, G., Giles Pacheco, F., Rodríguez-Vidal, J., Stringer, C.B. & Martinez Ruiz, F. 2008. Gorham's Cave, Gibraltar –the Persistence of a Neanderthal Population. *Quaternary International,* 181: 64-71.

Higham, T., Jacobi, R. M. and Bronk Ramsey, C. 2006. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* 48(2), 179-95.

Higham, T., Brock, F., Peresani, M., Broglio, A., Wood, R. and Douka, K. 2009. Problems with radiocarbon dating the Middle to Upper Palaeolithic transition in Italy. *Quaternary Science Reviews* 28, 1257-67.

Higham, T.F.G., Jacobi, R., Julien, M., David, F., Basell, L., Wood, R., Davies, W. & Bronk Ramsey, C. 2010. Chronology of the Grotte du Renne (France) and implications for the context of ornaments and human remains within the Châtelperronian. *Proceedings of the National Academy of Sciences (USA)* 107, 20234-9.

Higham, T. F. G., Brock, F., Bronk Ramsey, C., Davies, W., Wood, R. and Basell, L. 2011a. Chronology of the site of Grotte du Renne, Arcy-sur-Cure, France: implications for Neanderthal symbolic behaviour. *Before Farming* 201/2, 1-9.

Higham, T. F. G., Brock, F., Bronk Ramsey, C., Davies, W., Wood, R. and Basell, L. 2011b. Chronology of the Grotte du Renne, Arcy-sur-Cure, France: a response to Zilhão et al. (this issue). *Before Farming* 2011/3, 1-5.

Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Trinkaus, E., Chandler, B., Gröning, F., Collins, C., Hillson, S., O'Higgins, P., FitzGerald, C. and Fagan, M. 2011c. The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479, 521-4.

Higham,, T., Jacobi, R., Basell, L., Bronk Ramsey, C., Chiotti, L., and Nespoulet, R. 2011d. Precision dating of the Palaeolithic: a new radiocarbon chronology for the Abri Pataud (France), a key Aurignacian sequence. *Journal of Human Evolution* 61, 549-63.

Higham, T., Basell, L., Jacobi, R., Wood, R., Bronk Ramsey, C. and Conard, N. J. 2012. Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geissenklösterle. *Journal of Human Evolution* 62, 664-76.

Higham, T., Douka, K., Wood, R., Bronk Ramsey, C., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N. J., Draily, C., Froment, A., Galván, B., Gambassini, P., Garcia-Moreno, A., Grimaldi, S., Haesaerts, P., Holt, B., Iriarte-Chiapusso, M.-J., Jelinek, A., Jordá Pardo, J. F., Maíllo-Fernández, J. M., Marom, A., Maroto, J., Menéndez, M., Metz, L., Morin, E., Moroni, A., Negrino, F., Panagopoulou, E., Peresani, M., Pirson, S., de la Rasilla, M., Riel-Salvatore, J., Ronchitelli, A., Santamaria, D., Semal, P., Slimak, S., Soler, J., Soler, N., Villaluenga, A., Pinhasi, R. and Jacobi, R. 2014. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512, 306-9.

Hublin, J.-J., Talamo, S., Julien, M., David, F., Connet, N., Bodu, P., Vendermeersch, B. and Richards, M. P. 2012. Radiocarbon dates from the Grotte du Renne and Saint-Césaire support a Neanderthal origin for the Châtelperronian. *Proceedings of the National Academy of Sciences (USA)*109(46), 18743-8.

Jöris, O. and Street, M. 2008. At the end of the ^{14}C time scale – the Middle to Upper Palaeolithic record of western Eurasia. *Journal of Human Evolution* 55, 782-802.

Killick, D. 2015. The awkward adolescence of archaeological science. *Journal of Archaeological Science* 56, 242-7.

McPherron, S., Talamo, S., Goldberg, P. Niven, L., Sandgathe, D., Richards, M. P., Richter, D., Turq, A. and Dibble, H. L. 2012. Radiocarbon dates for the late Middle Palaeolithic at Pech de l'Azé, France. *Journal of Archaeological Science* 39, 3436-42.

Mellars, P. 1999. The Neanderthal Problem Continued. *Current Anthropology,* 40: 341-64.

Mellars, P. 2004. Neanderthals and the Modern Human Colonization of Europe. *Nature,* 432: 461-5.

Mellars, P., Gravina, B. and Bronk Ramsey, C. 2007. Confirmation of Neanderthal/Modern Human Interstratifications at the Châtelperronian Type-site. *Proceedings of the National Academy of Sciences (USA),* 104: 3657-62.

Mellars, P. (2010) Neanderthal symbolism and ornament manufacture: the bursting of a bubble? *Proceedings of the National Academy of Sciences (USA)* 107(47), 20147-8.

Millard A. R. 2003. Taking Bayes Beyond Radiocarbon: Bayesian Approaches to Some Other Chronometric Methods. In Buck, C. E. and Millard, A. R. (eds.) *Tools for Constructing Chronologies: Crossing Disciplinary Boundaries*. London: Springer, 231-248.

Moroni, A., Boscato, P. and Ronchitelli, A. 2012. What roots for the Uluzzian? Modern behaviour in central-southern Italy and hypotheses on AMS dispersal routes. *Quaternary International* 316, 27-44.

Pettitt, P.B., Davies, S.W.G., Gamble, C.S. & Richards, M.B. 2003. Palaeolithic Radiocarbon Chronology: Quantifying Our Confidence Beyond Two Half-lives. *Journal of Archaeological Science,* 30**:** 1685-1693.

Pirson, S., Flas, D., Abrams, G., Bonjean, D., Court-Picon, M., di Modica, K., Draily, C., Damblon, F., Haesaerts, P., Miller, R., Rougier, H., Toussaint, M. and Semal, P. 2012. Chronostratigraphic context of the Middle to Upper Palaeolithic transition: recent data from Belgium. *Quaternary International* 259, 78-94.

Reimer, P. J., Baillie, M. G. L., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Bronk Ramsey, C., Buck, C. E., Burr, G. S., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., McCormac, F. G., Manning, S. W., Reimer, R. W., Richards, D. A., Southon, J. R., Talamo, S., Turney, C. S. M ., van der Plicht, J. and Weyhenmeyer C. E. 2009. IntCal09 and marine09 radiocarbon age calibration curves, 0-50,000 years cal BP. *Radiocarbon* 51(4), 1111-50.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Bronk Ramsey, C., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., and van der Plicht, J. 2013. IntCal13 and Marine13 radicoarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* 55(4), 1869-87.

Riede, F. and Edinborough, K. 2012. Bayesian radiocarbon models for the cultural transition during the Allerød in southern Scandinavia. *Journal of Archaeological Science* 39, 744-56.

Roebroeks, W. 2008. Time for the Middle to Upper Paleolithic Transition in Europe. *Journal of Human Evolution,* 55: 918-26.

Steier, P. and Rom, W. 2000. The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42(2), 183-98.

Steier, P., Rom, W. and Puchegger, S. 2001. New methods and critical aspects in Bayesian mathematics for ^{14}C calibration.

Talamo, S. and Richards, M. 2011. A comparison of bone pretreatment methods for AMS dating of samples >30,000 BP. *Radiocarbon* 53(3), 443-9).

Talamo, S., Soressi, M., Roussel, M., Richards, M. and Hublin, J.-J. 2012a. A radiocarbon chronology for the complete Middle to Upper Palaeolithic transitional sequence of Les Cottés (France). *Journal of Archaeological Science* 39, 175-83.

Talamo, S., Hughen, K. A., Kromer, B. and Reimer, P. J. 2012b. Debates over Palaeolithic chronology – the reliability of ^{14}C is confirmed. *Journal of Archaeological Science* 39, 2464-7.

Teyssandier, N. 2003. *Les débuts de l'Aurignacien en Europe. Discussion à partir des sites de Geissenklösterle, Willendorf II, Krems-Hundsteig et Bacho Kiro.* PhD Dissertation, University of Paris X.

Tzedakis, P.C., Hughen, K.A., Cacho, I. & Harvati, K. 2007. Placing Late Neanderthals in a Climatic Context. *Nature,* 449: 206-8.

Unwin, S. 2003. *The Probability of God. A Simple Calculation that proves the Ultimate Truth.* New York: Three Rivers.

Vaquero, M. and Carbonell, E. 2012. Some clarifications on the Middle-Upper Palaeolithic transition in Abric Romaní: reply to Camps and Higham (2012). *Journal of Human Evolution* 63, 711-17.

Verpoorte, A. 2005. The first modern humans in Europe? A closer look at the dating evidence from the Swabian Jura (Germany). *Antiquity* 79, 269-79.

Walker, M.J., Gibert, J., López, M.V., Lombardi, A.V., Pérez-Pérez, A., Zapata, J., Ortega, J., Higham, T., Pike, A., Schwenninger, J-L., Zilhão, J. & Trinkaus, E. 2008. Late Neanderthals in Southeast Iberia: Sima de las Palomas del Cabezo Gordo, Murcia, Spain. *Proceedings of the National Academy of Sciences (USA),* 105, 20631-6.

Weninger, F., Steier, P., Kutschera, W. and Wild, E. M. 2010. Robust radiocarbon analysis, an attempt to improve Bayesian sequencing. *Radiocarbon* 52(2-3), 962-83.

White, M. J. and Pettitt, P. B. 2012. Ancient digs and modern myths. The context and age of the Kent's Cavern 4 maxilla and the spread of *Homo sapiens* in Europe. *European Journal of Archaeology* 15, 1-30.

Wood, R. 2015. From revolution to convention: the past, present and future of radiocarbon dating. *Journal of Archaeological Science* 56, 61-72. *Radiocarbon* 43(2A), 373-80.

Zilhão, J. 2006. Chronostratigraphy of the Middle to Upper Paleolithic Transition in the Iberian Peninsula. *Pyrenae,* 37(1), 7-84.

Zilhão, J. 2013. Neandertal-Modern Human Contact in Western Eurasia: Issues of Dating, Taxonomy, and Cultural Associations. in Akazawa, T., Nishiaki, Y. and Aoki, , K. (eds.) *Dynamics of Learning in Neanderthals and Modern Humans. Volume 1: Cultural Perspectives.* Replacement of Neanderthals by Modern Humans Series, Tokyo: Springer, 21-57.

Zilhão, J. & d'Errico, F. 1999. The Chronology and Taphonomy of the Earliest Aurignacian and its Implications for the Understanding of Neanderthal Extinction. *Journal of World Prehistory,* 13**:** 1-68.

Zilhão, J. & d'Errico, F. 2003a. The Chronology of the Aurignacian and Transitional Technocomplexes. Where do we stand? In*:* Zilhão, J. & d'Errico, F., eds, *The Chronology of the Aurignacian and of the Transitional Technocomplexes. Dating, Stratigraphies, Cultural Implications.* Trabalhos de Arqueologia, 313-349.

Zilhão, J. & d'Errico, F. 2003b. An Aurignacian 'Garden of Eden' in southern Germany? An alternative interpretation of the Geissenklösterle and a critique of the Kulturpumpe model. *Paleo* 15, 69-86.

Zilhão, J. and Pettitt, P. 2006. On the New Dates for Gorham's Cave and the Late Survival of Iberian Neanderthals. *Before Farming,* 2006/3: 95-122.

Zilhão, J., d'Errico, F., Bordes, J-G., Lenoble, A., Texier, J-P. and Rigaud, J-Ph. 2006. Analysis of Aurignacian Interstratifications at the Châtelperronian Type Site and Implications for the Behavioural Modernity of Neanderthals. *Proceedings of the National Academy of Sciences (USA),* 103: 12643-8.

Zilhão, J. D'errico, F., Bordes, J.-G., Lenoble, A., Texier, J.-P. and Rigaud, J.-Ph. 2008. Grotte des Fées (Châtelperron): History of Research, Stratigraphy, Dating, and Archaeology of the Châtelperronian Type-Site. *Paleoanthropology*, 1-42.

Zilhão, J. D'errico, F., Bordes, J.-G., Lenoble, A., Texier, J.-P. and Rigaud, J.-Ph. 2008. Like Hobbes' Chimney Birds. *Paleoanthropology*, 65-67.

Zilhão, J., Davis, S.J.M., Duarte, C., Soares, A.M.M., Steier, P. and Wild, E. 2010. Pego do Diabo (Loures, Portugal): Dating the Emergence of Anatomical Modernity in Westernmost Eurasia. *PLoS ONE,* 5(1): e8880.

Zilhão, J., d'Errico, F., Julien, M. and David, F. 2011a. Chronology of the site of Grotte du Renne, Arcy-sur-Cure, France: implications for radiocarbon dating. *Before Farming* 2011/3, 1-14).

Zilhão, J., Cardoso, J.L., Pike, A.W.G. and Weninger, B. 2011b. Gruta Nova da Columbeira (Bombarral, Portugal): Site Stratigraphy, Age of the Mousterian Sequence, and Implications for the Timing of Neanderthal Extinction in Iberia. *Quartär* 58, 93-112.