Seeking the Holy Grail: robust chronologies from archaeology and radiocarbon dating combined

Eszter Bánffy,1 Alex Bayliss,2 Anthony Denaire,3 Bisserka Gaydarska,4 Daniela Hofmann,5 Philippe Lefranc,6 János Jakucs,7 Miroslav Marić,8 Krisztián Oross,7 Nenad Tasić,9 and Alasdair Whittle4

1 Römisch-Germanische Kommission, Frankfurt am Main, DE
eszter.banffy@dainst.de
2 Historic England, London, UK, and Biological and Environmental Sciences, University of Stirling, Stirling, UK
alex.bayliss@historicengland.org.uk
3 Department of Archaeology and Art History, and UMR 6298 Artehis, University of Burgundy, Dijon, F
Anthony.Denaire@u-bourgogne.fr
4 Department of Archaeology and Conservation, Cardiff University, Cardiff, UK
b_gaydarska@yahoo.co.uk; whittle@cardiff.ac.uk
5 Department of Archaeology, Bergen University, Bergen, NO
Daniela.Hofmann@uib.no
6 Institut national de recherches archéologiques préventives, and UMR 7044 Archimède, University of Strasbourg, Strasbourg, FR
philippe.lefranc@inrap.fr
7 Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest, H
jakucs86@gmail.com; oross.krisztian@btk.mta.hu
8 The Institute for Balkan Studies, Serbian Academy of Science and Arts, Belgrade, RS
dzibril@gmail.com
9 Faculty of Philosophy, Belgrade University, Belgrade, RS
ntasic@gmail.com

ABSTRACT – The strengths of formal Bayesian chronological modelling are restated, combining as it does knowledge of the archaeology with the radiocarbon dating of carefully chosen samples of known taphonomy in association with diagnostic material culture. The risks of dating bone samples are reviewed, along with a brief history of the development of approaches to the radiocarbon dating of bone. In reply to Strien (2017), selected topics concerned with the emergence and aftermath of the LBK are discussed, as well as the early Vinča, Ražište and Hinkelstein sequences. The need for rigour in an approach which combines archaeology and radiocarbon dating is underlined.

KEY WORDS – radiocarbon dating; bone samples; laboratory methods; Bayesian chronological modelling; earliest LBK; Vinča; Ražište; Hinkelstein

IZVLEČEK – Utrjujemo moč formalnega Bayesovega kronološkega modeliranja z združevanjem vede
nja iz arheologije z radiokarbonskim datiranjem skrbno izbranih vzorcev znane tafonomije in dia-
gnostične materialne kulture. Ponorno preučimo nevarnosti datiranja kostnih vzorcev skupaj s krat-
kim pregledom razvoja pristopov k radiokarbonskemu datiranju kosti. Razpravljamo tudi o izbra-
nih temah, povezanih s pojavom in posledico LTK ter sekvencami zgodnje Vinče, Ražišče in Hinkel-
steina kot odgovor na Striena (2017). Poudarjamo tudi potrebo po natančnosti/doslednosti pri zdru-
ževanju arheologije in radiokarbonskega datiranja.

IZVLEČKA – radiokarbonsko datiranje; vzorci kosti; laboratorijske metode; Bayesovo krono-
loško modeliranje; najzgodnejša LTK; Vinča; Ražište; Hinkelstein

KLJUČNE BESEDE – radiokarbonsko datiranje; vzorci kosti; laboratorijske metode; Bayesovo krono-
loško modeliranje; najzgodnejša LTK; Vinča; Ražište; Hinkelstein
Introduction

Recently in this journal, Hans-Christoph Strien (2017) suggested that there are discrepancies between archaeological and 14C-based chronologies, particularly in relation to a series of new studies considering the early and middle Neolithic sequences across a swathe of Europe from Serbia in the south-east to the lower Rhineland in the north-west (Jakucs et al. 2016; Oross et al. 2016a–c; Tasić et al. 2016a; 2016b; Denaire et al. 2017). He ended his paper by stating that “radiocarbon dating is not the Holy Grail of prehistoric archaeology, especially as long as environmental influences on its results are neither fully recognised nor understood” (Strien 2017: 279).

We consider this critique to be problematic on a number of levels, the most fundamental being the implicit oppositional dualism of archaeology and radiocarbon dating. All these new studies have formal statistical modelling of archaeological chronologies at their heart, and employ explicit Bayesian methodologies to weave together the available archaeological information with the scientific dating evidence (Bayliss, Whittle 2015; 2018). This is a holistic approach, and so there can be no opposition between archaeological and 14C-based chronologies. The choice is between archaeological chronologies that cross-reference to each other and make only informal reference to scientific dating evidence, and modelled chronologies which formally combine all the different strands of evidence. We restate the strengths of a formal approach.

Before investigating certain aspects of Strien’s approach, it is important to stress that there is, in fact, a long tradition in continental European research of combining radiocarbon dating and archaeological information (see, for example, Breunig 1987; Weninger 1995; Müller 2009), although so far, few studies have undertaken a detailed critique of the character of the dated material and its archaeological associations that we consider essential for constructing robust chronologies (Bayliss et al. 2011; 2016). But the study of both forms of evidence together is certainly a vigorous strand in this tradition.

Strien’s critique implies that “archaeology” is led principally by chronotypology, whereas we argue for a much wider and more inclusive set of evidence, including understandings of context, stratigraphy, taphonomy and associations; but importantly, we agree on the key value of detailed knowledge of the material. Although it is not fully explicit in the article (Strien 2017), it is also relevant to note some of the key assumptions and approaches underlying Strien’s vision of the earliest Neolithic, specifically the earliest LBK, in central Europe, which he has set out at greater length elsewhere (Strien 2017). We will deal with selected aspects of this perspective below, but note by way of introduction an apparently sceptical if not hostile attitude to radiocarbon dating in general, and the accuracy of measurements on bone samples in particular (Strien 2018: 27–28, 65–66). This is combined with the belief, based on informal inspection of selected radiocarbon dates, reinforced in part by reliance on a very speculative estimate for the well at Mohelnice in Moravia, that the earliest LBK goes back to towards 5600 cal BC (Strien 2018: 28). A belief is also evident that chronotypology, especially through detailed study of arrangements of decorative motifs on pottery, can serve to outline a succession of so-called house generations, with equal validity across a wide swathe of central Europe (Strien 2018: 32, Abb. B5–B12). And there is an inherent tendency to prefer continuity in material development over possibilities of interruption, gap or hiatus. Since two of the previous studies (Tasić et al. 2016a; Denaire et al. 2017; see also Tasić et al. 2016b) are regarded as generally unproblematic by Strien (2017: 272–273), it seems clear that reservations and doubts about the results of the third study (Jakucs et al. 2016) are driven by other factors than just scepticism about radiocarbon dating in general, including a desire to defend an alternative, “higher” chronology for the earliest LBK.

Risk and radiocarbon dating

In paradise, all radiocarbon samples date the target event intended and all radiocarbon measurements are accurate to within their quoted uncertainty. The real world is not like this. Few radiocarbon samples, and even fewer sampling strategies, are perfect. There is always some element of risk in dating a group of samples. These risks are of two kinds – archaeological risks and scientific risks – both of which must be managed during a programme of radiocarbon dating.

The major archaeological risk is the association between the dated event and the target event (Waterbolk 1971). Except in rare cases where the item dated is itself the topic of interest (e.g., a carbonised food crust on a pottery sherd with diagnostic decorative...
motifs), this relationship is never known, but is inferred on the basis of archaeological evidence. This is why stringent archaeological criteria have been developed to assess the security of the relationships between potential samples and the contexts from which they were recovered (Bayliss et al. 2011; 2016). It is also possible to mitigate the risk of residual or intrusive samples by selecting single-entity samples for dating (Ashmore 1999), or dating multiple fragments from a single deposit (Bayliss et al. 2014).

Scientific risks are more varied. They include the potential for age-at-death offsets and reservoir effects, heterogeneity in bulk samples, diagenesis (stressed especially by Strien 2017) and contamination of sampled materials, and laboratory inaccuracy. Again, these risks should be identified and managed. Some can be avoided by, for example, obtaining botanical identifications of charred plant material and selecting only short-lived, single fragments for dating. Others cannot always be avoided, and so mitigation strategies are required. Where models rely heavily on samples of human bone, for example, the potential for reservoir offsets related to human diet may be investigated through stable isotopic analyses (e.g., Fernandes et al. 2014) or the dating of “perfect pairs” of contemporaneous human and herbivore bone (e.g., Bayliss et al. 2016; Fig. 6). Diagenesis and contamination of sampled materials and potential laboratory inaccuracy are also issues which cannot be avoided, and so have to be mitigated. Radiocarbon dating of archaeological materials involves a variety of complex processes, which have been, and remain, an active area of scientific research. Methods therefore develop over time, and it is essential that interpretations of radiocarbon measurements take full and proper account of these methodological developments.

Laboratories themselves take the issue of measurement accuracy extremely seriously, employing the international standard reference material, Oxalic Acid II (Mann 1983), background standards that are devoid of 14C (e.g., van der Plicht et al. 2000; Fig. 5), and a range of secondary standards which are dated repeatedly, both as a check to identify when something may have gone wrong in processing a particular batch of samples, and to determine over the long-term how the actual scatter of results compares with those expected on the basis of the quoted errors (e.g., Staff et al. 2014; Fig. 1). Over the past 30 years, a series of formal international inter-comparison exercises have been undertaken (e.g., Scott et al. 2017), and recently there have also been several smaller inter-comparison exercises specialising in specific material types (e.g., Naysmith et al. 2007). These procedures can, and do, identify problems and allow them to be eliminated (e.g., Bronk Ramsey et al. 2002.2).

The accuracy of a suite of radiocarbon dates can also be assessed once they have been obtained, both individually and as a group. There are a number of methods that we can use as a check on our results:

- the consistency of replicate results on the same or similar material (see Ward, Wilson 1978);
- the coherence of a suite of related radiocarbon dates – are there any clear outliers or misfits (see Bayliss et al. 2016.56)?
- and the compatibility of a series of results with the relative chronological sequence known from archaeological information (such as stratigraphy) (e.g., Bronk Ramsey 2009a; 2009b).

Alex Bayliss and Peter Marshall (submitted) have recently reviewed groups of replicate measurements on 1089 archaeological samples. They found that overall approx. 12% of results lay more than 2σ from the true value (rather than the 5% expected on statistical grounds alone). Some materials are clearly more problematic than others (Tab. 1). For example, approx. 30% of results on carbonised residues on the interior of pottery sherds are problematic, but replicate results on single-entity charred plant remains vary only according to statistical expectation (5% lie more than 2σ from the true value).

<table>
<thead>
<tr>
<th>Sample Material</th>
<th>Archaeological Risk</th>
<th>Scientific Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-1993 measurements</td>
<td>Variable</td>
<td>Medium</td>
</tr>
<tr>
<td>Sediments</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Carbonised residues</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Wood (multi-ring, mostly waterlogged)</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Wood (single-ring, mostly from buildings)</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Single-entity charred plants</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Waterlogged plants</td>
<td>Variable</td>
<td>Low</td>
</tr>
<tr>
<td>Bone &amp; antler</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Calcined bone</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

Tab. 1. Risks in radiocarbon dating of different archaeological sample types. Archaeological risks have been assessed informally following Bayliss et al. (2011.56–58); scientific risks have been quantified by Bayliss and Marshall (submitted: High, more than 20% outside 2σ; Medium, 10–20% outside 2σ; Low, less than 10% outside 2σ).
The accuracy of radiocarbon measurements on bone and antler samples

Such considerations should allay fears about, and suspicion of, the reliability of radiocarbon dating in general, but they underline the need for vigilance and rigour throughout the dating process.

The replicate analysis summarised in Table 1 does not suggest that samples of bone and antler are particularly problematic but given the concerns raised about the accurate dating of this material type by Strien (2017), it is worth delving into the evidence in a little more depth.

Of measurements on bone or antler made before 1993 and replicated randomly, some 20% appear to lie more than 2σ from the true value. Of conventional measurements made after 1993 and similarly replicated randomly, approx. 12% lie outside the 2σ limits. These findings reflect the pretreatment protocols available for conventional dating and the difficulty of providing sufficient material for this process. This is exemplified by the series of radiocarbon dates on human bone from the Trebur cemetery undertaken at the Heidelberg laboratory in the 1980s, which appear to be anomalously recent (Spätz 2001). This probably results from poorly preserved bone with low collagen yields, where diagenetic alterations include the attachment of exogenous humic materials to the protein strands (Hedges, van Klinken 1992). Humic acids are generally, but not invariably, younger contaminants.

The advent of Accelerator Mass Spectrometry initially did little to improve this situation, as the new technology simply allowed results to be obtained on low-collagen samples that were previously undatable. This was particularly problematic when large samples were processed for conventional radiocarbon dating, but produced such low collagen yields that they had to be dated by AMS. This was the case, for example, with the series of animal bone samples prepared for liquid scintillation spectrometry at the University of Zürich, but dated by AMS at ETH Zürich (Irka Hadjas, pers. comm., October 2018) from the settlement at Rottenburg “Fröbelweg”. These results again appear anomalously recent for the associated älteste or earliest LBK ceramics (Bofinger 2005; Jacques et al. 2016). Early AMS dating of bone from Trebur produced mixed results (Bronk Ramsey et al. 2002.16–17), with measurements on compact femurs appearing more reliable than those on spongy vertebrae (Spätz 2001.283).

Considerable attention was paid to improving methods of bone pretreatment throughout the 1980s and 1990s (Hedges, van Klinken 1992; Tisnérat-Labarde et al. 2003). The first significant advance was the realisation that some bones are simply too poorly preserved for accurate radiocarbon dating, and that thresholds for minimum collagen yield (van Klinken 1999) and maximum C:N ratio (DeNiro 1985) should be employed. Although a wide variety of methods have been suggested for bone pretreatment, for samples of Holocene age from temperate climates, those commonly employed basically can be divided into variants of that outlined by Robert Longin (1971) and those that utilise ultrafiltration (Brown et al. 1988). In most circumstances, these produce comparable results. For example, approx. 8% of results on bone and antler samples in the 359 AMS replicate groups considered by Bayliss and Marshall (submitted) appear to lie more than 2σ from the true value. No significant difference is observed between the results from samples processed by both methods (the mean difference is 8.2±10.8 BP). Ultrafiltration may, however, be more effective when dating older samples (Higham et al. 2006; Talamo, Richards 2011).

This comparability of results is also observed with the radiocarbon measurements on bone samples from the studies considered by Strien (2017), all of which were processed using ultrafiltration. Figure 1 illustrates the differences between pairs of measurements on the same sample from these sites. Of the 28 pairs of results on bone samples, 25 are statistically consistent at the 5% significance level, and the mean difference is 11.1±13.3 BP (all six pairs of results on charred plant remains are statistically consistent at the 5% significance level, and the mean difference is 0.5±18.0 BP).

This inter-laboratory reproducibility gives us confidence in the reliability of the radiocarbon measurements on bone since, if the samples were too poorly preserved for accurate dating, it is extremely unlikely that the laboratory processes in two different facilities would fail to remove exactly the same proportion of contamination in each sample. The consistency of the results suggests that both laboratories have succeeded in removing exogenous carbon from the samples, and have dated carbon purely derived from the archaeological specimens. Inter-laboratory replication therefore provides one strategy for mitigating the scientific risks of sampling strategies that are highly dependent on one type of datable material (in this case, bone).
There are other approaches which we can use to validate the accuracy of a suite of radiocarbon dates. One of these is the compatibility of the dates with the prior information included in a Bayesian model, which is particularly powerful when the archaeological evidence provides a secure relative sequence. An example of this approach is provided by the 101 radiocarbon measurements on samples taken from the approx. 7 metre-deep section through the tell of Vinča-Belo Brdo excavated in 2004–2005 and 2012–2014 (Tasić et al. 2016b.Fig. 10). Ninety-three of these results are included in the chronological model described by Nenad Tasić et al. (2016b.Figs. 3–8), with eight results on short-lived, single-entity charred plant remains excluded, as the samples were clearly intrusive in the contexts from which they were recovered. Thirteen further results are included only as termini post quos for overlying deposits; one sample of calcined bone appears to have incorporated a component of old wood during the cremation process (Olsen et al. 2013; Snoeck et al. 2014) and 12 samples appear to be residual (two disarticulated animal bones and ten single-entity samples of charred plant material). The dates on all 11 animal bone samples included in this model have good individual agreement ($A > 60$; Bronk Ramsey 1995.426), and the model itself has good overall agreement (Amodel: 72; Bronk Ramsey 2009a.356–357). These statistical indicators show that the dates on the bone samples are compatible both with the dates on other materials included in the model (overwhelmingly short-lived charred plant remains), and with the relative sequence of deposits known from stratigraphy.

This application also illustrates the balance between different kinds of risk that must be assessed in constructing a sampling strategy for radiocarbon dating. Bones require complex chemistry in the laboratory, but when articulating or refitting specimens are selected for dating (Bayliss et al. 2016.Fig. 7), the archaeological risks of intrusion or residuality are very low. The scientific risk is therefore mitigated by dating a selection of specimens at two facilities. Charred plant remains are generally easier to process in the laboratory, but are much more likely to be intrusive or residual. At Vinča, where this risk can be quantified by reference to the stratigraphic sequence, 9% of single, short-lived charred plant remains are intrusive and 11% residual. This combined archaeological risk is mitigated by dating two separate single-entity samples of short-lived charred plant remains from every deposit. Residual and intrusive samples cannot be avoided, but they can be identified and modelled appropriately. This requires replication, which is clearly an essential tool in mitigating risk in radiocarbon dating.

The accuracy of radiocarbon dates on bone samples is also validated by the quality assurance procedures undertaken by laboratories. Typically, this involves the repeat preparation and dating of a Pleistocene bone sample that is known to be so old that all the radiocarbon within it has decayed away. Any mea-
sured radiocarbon within it thus represents either natural contamination which has not been adequately removed from the sample or carbon introduced during laboratory processing. These “processed blanks” quantify any possibility that radiocarbon may be present in a bone sample that is not derived from the bone itself (e.g., Brock et al. 2007, Fig. 3).

Many laboratories also undertake repeat preparation and dating of a more recent animal bone sample, often one that is of known age. These bone standards provide a check that the radiocarbon content of a sample has not been diluted by the introduction of radiocarbon-free carbon of petro-chemical origin in the laboratory (e.g., Brock et al. 2007, Fig. 2). The radiocarbon content of the known-age bone sample can also be compared to that of contemporaneous wood samples whose date is known from dendrochronology, thus providing a direct check on the calendar accuracy of the bone date.

Finally, the accuracy of radiocarbon dates can be compared with calendar dates from equivalent contexts provided by dendrochronology (cf. Ullrich 2008, 73–79). We are, frankly, puzzled by the contention that there is any inconsistency between the typological evidence, the radiocarbon dating of the bone samples considered in this paper, and the available tree-ring dates. Figure 2 shows the available tree-ring dates from LBK contexts that can be related typologically to the dated sequence from Lower Alsace (we note the possibility that changes between ceramic phases may not be exactly synchronous in different LBK groups). The first well at Leipzig-Plaussig is not related to any diagnostic sherds, but must date to after the settlement was founded in LBK II. The second well here is later than the first, and associated with pottery that is equivalent to phase IV in the Alsace series (Friederich 2017). The first well at Erkelenz-Kückhoven must date to after the beginning of LBK II (as there is no LBK I in this area), and is earlier than the second well, which is associated with pottery whose decoration can be paralleled with phase IVb in Alsace (Weiner 1998). The use of another well at Altscherbitz in eastern Germany is associated with younger LBK pottery and Šárka ceramics, and can be tentatively equated with phases IVb or V in Alsace (Tegel et al. 2012). In all cases, the tree-ring dating is clearly compatible with the chronological modelling of the Alsace sequence. In this light, it is the very tentative cross-dating of the extremely short tree-ring series from Mohelnice, Moravia, that stands out as anomalous, and we agree with Michael Friedrich that it should not be regarded as absolutely dated (Schmidt, Grühle 2003, 58; Friederich 2017, 430–431; see also Cladders 2018, 4).

Fig. 2. Comparison between tree-ring dates for features containing diagnostic assemblages of LBK pottery and posterior date estimates for equivalent ceramic phases in Lower Alsace, derived from the model presented by Denaire et al. (2017, Fig. 8). Modelled parameters are given in italics.

1 Several wells were found at the Mohelnice “Volutenkeramik” site (Tichý 1972), of which only that numbered CXXXIV is relevant here, because Tichý regarded it as an older LBK feature. The timber fragments were very fragmented, as they were found in the waterlogged soil of the well. They were kept wet, and even deep frozen, in order to prevent further damage, while being delivered to Cologne for tree-ring analysis. Out of seven small wooden pieces, only two were suitable for dating, but given the relatively small number of tree-rings, secure tree-ring dating of the well was not possible (Schmidt, Grühle 2003, 56). Two samples from this same well have also been radiocarbon dated: GrN-6610: 6240±65 BP and KN-4339: 6580±75 BP; Schmidt and Grühle (2003, 56) hazarded the opinion that the earlier result might be “closer to the real date of the well” (our translation). Strien (2017, 277) claims, in our view without sufficient support, that the well in question is “not later than 5400 BC.”
in Alsace can be associated with dendrochronology. In the wetland settlement of Egolzwil 3, a cultural layer containing local pottery along with two imported vessels with Bruebach-Oberbergen decoration is dated to between 4282 BC and 4274 BC (Denaire et al. 2011). Two wells from Dambach-la-Ville in Lower Alsace contained assemblages of Bischheim Occidental du Rhin Supérieur I (BORS I) pottery (Croutsch et al. 2016). Again, in these cases, the tree-ring dates are clearly compatible with the chronological modelling (Fig. 3).

Other issues: classifications and concepts

On this basis, we stand by our published papers (Jakucs et al. 2016; Oross et al. 2016c; Tasić et al. 2016a; 2016b; Denaire et al. 2017). There is no space here to deal with each and every individual site raised by Hans-Christoph Strien (2017). From his list, we have discussed Mohelnice, Rottenburg2 and Trebur above, and it is also the case that the radiocarbon dating so far of samples from Herxheim (Strien 2017.277) has been hampered by very poor collagen preservation in bone from the site (see Riedhammer forthcoming).

Two further cases do, however, deserve a little more comment. First, it should be noted that the human bone from the “founder” burial at Schwanfeld which gave a very early date (Hd-14219; Feature 704/760: 6580±20 BP; see Stäuble 2005) has now been redated (Fröhlich, Lüning 2017.43–49; Tab. 1). The three new AMS measurements on this male skeleton are not statistically consistent at the 5% significance level, although they are at the 1% significance level (OxA-25035, 6300±40 BP; OxA-26143, 6351±37 BP; MAMS-15659, 6228±26 BP; T' = 7.9, T(5%) = 6.0, T'(1%) = 9.2, v = 2), and are all considerably younger than the original measurement, which appears to have been contaminated by Paraloid B72. We have combined the sub-set of radiocarbon dates from Schwanfeld listed by Fröhlich and Lüning (2017. Tab 1) with the sequence of house generations suggested by them. This model has good overall agreement (Amodel: 85; Fig. 4), and the weighted mean (6276±19 BP) of the AMS results on the “founder” grave has good individual agreement (A: 101). All three conventional measurements on bone in this model, however, have very low probabilities of falling in the house generation suggested (P: 2, P: 0, P:2 respectively; Fig. 4). This suggests that their accuracy cannot be relied upon. This means that the model contains only four fully effective likelihoods (two on bone samples dated recently by AMS, one on a charred cereal grain, and one on a carbonised residue on a pottery sherd). Consequently, the date estimate for the start of occupation at Schwanfeld, which occurred in 5680–5220 cal BC (95% probability; start Schwanfeld; Fig. 4), probably in 5420–5245 cal BC (68% probability), is insufficiently precise to contribute meaningfully to this debate. It should be noted, however, that the limited number of reliable radiocarbon dates currently available are compatible both with the sequence of house generations suggested for Schwanfeld (Fröhlich, Lüning 2017) and with the first appearance of the LBK in this region in the mid-5th century cal BC (start west; Jakucs et al. 2016. Fig. 23).

Secondly, although the Vinča – Belo Brdo chronology is deemed to “pose no obvious problems” (Strien 2017.272), there are persistent attempts to under-
Seeking the Holy Grail: robust chronologies from archaeology and radiocarbon dating combined

127

mine the wider implications of the formally modelled chronologies. The first one concerns the start of Vinča A at the eponymous site and at Szederkeny. Unsubstantiated claims for a ground-water offset affecting the lowest levels of tells are aired, and the possibility of a “too late” start to Vinča A is mooted (Strien 2017:277). Interestingly, there is no comment on such possible effects on the samples from the earlier Starčevo samples from Belo Brdo. Should the hydrological effects have been unkind to the Vinča A samples only and not to Starčevo samples3, then there should be either an overlap between the periods at Belo Brdo, or the gap between the periods should be much shorter. The presence of a gap has been shown in two places at Belo Brdo: at the base of the deep sounding excavated in 2004–2005 and 2012–2014, where radiocarbon dates were obtained on short-lived charred plant material (Tasić et al. 2016b.Tab. 1), and at the base of the sequence excavated by Miloje Vasić, where radiocarbon dates were obtained on bone samples (Tasić et al. 2016a.Tab. 2). In the latter case, its duration has been estimated as 45–220 years (95% probability; gap; Tasić et al. 2016a.Fig. 18), probably 120–200 years (68% probability). It therefore seems unlikely that local hydrological conditions affected a range of different materials from only Vinča deposits at Belo Brdo, and thus the gap between the Starčevo and Vinča occupation at these two different locations should be considered as genuine, and the beginning of Vinča A

3 The earliest Vinča phase at Belo Brdo lies just above the prehistoric humus, which is of chernozem type. The Starčevo phase contexts are cut down from the chernozem itself on to the loess-like sediments below. One of the main characteristics of the chernozem soil is its porosity, so it is highly unlikely that either early Vinča or Starčevo samples would have been waterlogged long enough to influence their diagenesis. Furthermore, the existence of loess-like sediments immediately below would draw the water even deeper down.
ša A at Belo Brdo put in the first quarter of the 53rd century cal BC (Tasić et al. 2016a. Fig. 17).4

We agree with Strien, however, that vigilance and rigour need to be exercised in each and every case, and as we stressed in our paper (Jakucs et al. 2016), the current quality and quantity of radiocarbon dating for the earliest LBK as a whole undoubtedly have room for improvement. Nonetheless, we see no basis from the evidence currently available for as “high” or early a chronology for the emergence of the LBK as proposed by Strien, back towards 5600 cal BC. This matter seems to us also to raise other questions both of classification and conceptualisation, which we now briefly address.

Questions of classification

The formative LBK

There is a clash of classification in how to order and group material, especially pottery, at the start of the LBK. The “formative phase” was proposed in order to distinguish the beginnings from the “älteste”, or earliest, LBK identified in the German literature. It is based, on the one hand, on the Starčevo presence in southern Transdanubia and the Balaton area, ending perhaps in the 56th century cal BC (Oross et al. 2016a. Fig. 8). On the other hand, two longhouses in a style otherwise very typical in the earliest LBK were excavated at the site of Szentgyörgyvölgy-Pityerdomb west of Lake Balaton, associated with a pottery assemblage with strong Starčevo affinities (Bánffy 2000; 2004). All the features and the pottery coming from the individual features, some 15 000 sherds in total, have been described in detail, and there are now dozens of other comparable sites in the region suggested by small excavations or surface finds. Although all the usual domesticated plants and animals were present at Szentgyörgyvölgy-Pityerdomb, these small sites suggest scattered communities, not yet settled on loess soils and perhaps not fully dependent on agriculture (Kreuz 1990). The Szentgyörgyvölgy-Pityerdomb material (both its pottery and flints) is closely related to site Ila at Brunn near Vienna (Stadler 1995; Stadler, Kološa 2010). There is currently no clear modelled evidence for any non-formative LBK sites in Transdanubia predating the middle of the 54th century cal BC (Oross et al. 2016b. Fig. 9; Jakucs et al. 2016). Following settlement of the loess plateaus of the River Marcal and elsewhere in the western Carpathian basin, a rather rapid expansion, creating the spread of the LBK, appears to have begun. The speed and the nature of this spread have been attested by both mtDNA and whole genomic DNA analyses (Szécsényi et al. 2015; Lipson et al. 2017). Archaeologically, one of the best examples of the scenario is the case of Vedrovice in Moravia, where the first migrant generations can be distinguished from the later ones, which appear to have encountered and mingled with local groups (Zvelebil, Pettitt 2008).

In contrast, Strien simply reduces the proposed “formative phase” of the LBK to a regional variant, creating a “Balaton aLBK” group, separated from a “Danube aLBK” group (Strien 2017.273, 278–279; 2018.35–40). It is not clear how his own proposed start date for the earliest LBK of around 5600 cal BC is derived, other than by selective use of visual inspection of radiocarbon dates (arbitrarily switching between calibrated 14C dates and uncalibrated 14C determinations), aided and abetted by acceptance of a speculatively early date for the Mohelnice well (as discussed above); the issue is further muddied by claims that a much earlier start date than we have proposed would solve alleged problems with the demography of expansion. We reflect on that issue below.

Sopot and Vinča

The second non-explicit criticism of Strien (2017) of the Vinča-Belo Brdo chronology concerns more broadly the formal chronological models for the use and development of Vinča ceramics, namely the relationship between different ceramic styles like Sopot and Vinča. It is claimed (Strien 2017.276) that, allegedly on typological grounds, Sopot IB-II is related to Vinča C1 (Schier phase 6 at Vinča – Belo Brdo; Oross et al. 2016c.158–159), while radiocarbon dating relates it to Vinča A1 at Szederkény. Both of these statements are incorrect.

---
4 In relation to the start of Vinča A at Szederkény, the subdivision of the major Vinča phases is based on statistical analysis of pottery from the layers of the Vinča-Belo Brdo tell (Schier 1995; 1996). However, the sub-phases established on Belo Brdo are extremely difficult to reproduce in a large-scale flat settlement which is located at the northernmost edge of the Vinča distribution in the contact zone of multiple cultural units. In the very first publication of the Szederkény settlement (even without radiocarbon dating; Jakucs, Voicsek 2015), we proposed, on typological grounds, that the beginning of the site can be equated with the earliest Vinča phase (Vinča A1a, sensu Schier), although in general there are also comparisons with the whole of Vinča A (Jakucs, Voicsek 2015). Radiocarbon data later confirmed this first observation; and in the eastern part of the settlement, there are pots matching the Vinča A ceramic style as a whole (Jakucs et al. 2016).
First, Krisztián Oross et al. (2016c.158–159) do not simply equate Vinča C1 with Sopot IB-II, but followed the development of scholarly opinion on the chronology of the Sopot (initially Sopot-Bicske) distribution in Hungary. The entity was regarded as coeval with Sopot Ib and II by Nándor Kalicz and János Makkay at the time of its recognition, later connected more precisely to the Vinča B2–C horizon. As a result of the re-evaluation concerning the initial phase of the younger Vinča culture by Schier and inspired by his Vinča C1 phase (Schier 1996. 147–148), further Hungarian Sopot sites were dated similarly. Recent dating programmes for Vinča – Belo Brdo and for the Alsónyék Sopot occupation can only reinforce this context. Secondly, the particularities of Sopot ceramics in Croatia and Hungary, together with the largely unpublished early Sopot assemblages and the uncertainties of the radiocarbon dates (Balen et al. 2009.58), are entirely overlooked.

**Ražište, Sopot and Vinča**

It is, in fact, an anomaly that in the literature, Ražište sites (equivalent to Vinča A and B1 horizons) and Sopot sites (Vinča B2/C and Vinča C horizon) have both been labelled sometimes as Sopot IB-II. All the alleged contradictions discussed by Strien, in the section “Too young, but sometimes old: the case of Ražište/Sopot IB-II”, come mainly from the fact that he ignored this anomaly. The role of the Ražište-type in the biography of the Sopot culture is still debated today (Jakucs, Voicsek 2015; Jakucs et al. 2016.299). As Ražište-type ceramics are one of the main components of the Szederkény settlement, Jakucs et al. 2016 examined the original view that assigned the Ražište-type to the Sopot IA–IB or IB-II horizons (Marković 2012.58–59). Thus, we examined the coherence of the data specifically related either to the Ražište-type or to the Sopot IB-II (Jakucs et al. 2016.300, Tab. 4). This obviously should not have been an arbitrary selection, and so we could not omit the samples of Ivandvor-Šuma Gaj, since they were clearly listed as Sopot IB-II (Burić 2015), though they were apparently young. However, precisely for the above reasons, Jakucs et al. (2016) never mentioned Ivandvor-Šuma Gaj as a “Sopot IB-II/Ražište-site”, as Strien asserts in his paper (Strien 2017.273). We have also never tried to find a correlation between the Ražište finds of Szederkény and the Sopot finds of Alsónyék, as Strien suggested (Strien 2017.276), since these obviously differ typologically and cannot belong to the same Sopot horizon in any case. Once again, there appears to be no solid basis for arguing that radiocarbon dates are both too old and too young.

**Other conceptual issues: demography; and hiatus**

**Demography**

Strien asserts (2017.273) that accepting our modelled results “is not so much a chronological as a demographical problem”. In his view, a very rapid earliest LBK expansion would have entailed an improbably large number of people, with some 3000 people leaving the Balaton-Vienna area in the first year for south-west and central Germany, and thousands of immigrants between Bavaria and Volhynia, “an obviously unrealistic number” (Strien 2017.273). These demographic speculations are then used to cast doubt on the modelled 54th-century date for the major earliest LBK diaspora, with an informal estimate of “not later than 5500 cal BC” given instead (Strien 2017.273). This is hardly the place to go into the many difficulties involved in trying to reconstruct population levels in the earliest, or, indeed, the established LBK. Suffice it to say that there are so many uncertainties and imponderables involved, with classic questions of the numbers of occupants in longhouses, the numbers of longhouses in contemporaneous use in a given settlement, and the numbers and durations of individual settlements in any given region among many others (see Soudsky 1969; Moddermann 1970; Coudart 1998; Dubouloz 2008; Bocquet-Appel et al. 2014). We did not propose any specific figures, but considerable numbers in certain parts of the landscape are plausible in general (see also Shennan 2018), and the aDNA evidence now available (e.g., Lipson et al. 2017) is also compatible with an LBK diaspora of considerable size. Demography is hardly the kind of “deal-breaker” which Strien claims.

**Gaps and hiatus: the case of Hinkelstein**

In a last effort to cast doubt on the reliability of radiocarbon dating, Strien addresses the question of the date of the Hinkelstein phase, supporting the conventional view that there was unbroken continuity from LBK to Hinkelstein (Strien 2017.275, 278).

We can note, first, that for the early and middle Neolithic as a whole in Lower Alsace and by extension in a large part of the Upper Rhine valley, radiocarbon dating and archaeology (mainly typo-chronology and stratigraphy) have basically agreed in establishing the same sequence. Strien himself (2017.273) appears to accept the modelled results of Anthony Denaire et al. (2017) overall. The position of Hinkelstein in this sequence (Denaire et al. 2017...
1130–1137; see also Riedhammer in press) is the bone of contention. This takes us far from the earliest LBK, but is central to questions of dating as a whole.

In our view, the archaeological evidence for contacts between users of late LBK and of Hinkelstein pottery is limited to the middle Rhine valley, more specifically the Worms region and its Hinkelstein graveyards (Meier-Arendt 1975; Jeunesse 1999; Spatz 2001). Outside this region, there is no definite evidence of contact, because it is impossible to rely on the usually cited “mixed assemblages” containing both LBK and Hinkelstein finds as closed contexts, except that from Köln-Lindenthal (Buttler 1935; Spatz 2001); and in the latter case, we should underline that Köln is located outside the normal distribution of Hinkelstein pottery. Nothing that we are aware of definitely proves that the Hinkelstein style had emerged when the LBK ended in Lower Alsace or other regions, and the map distributions of late LBK and Hinkelstein respectively are also no proof of direct contact (see also Ritter-Burkert in press, on the situation in the Wetterau; and Riedhammer 2018.69).

We believe that the existence of a gap between the early and middle Neolithic should not be a surprise. In the Rhineband, a similar situation exists in the region of the Aldenhovener Platte, where the middle Neolithic sequence categorically starts with Grossgartach – recent Grossgartach in the Spatz chronology, though earlier in the Denaire system (Spatz 1996; Denaire 2009). The existence there of a gap between LBK and Grossgartach appears to be supported by pollen diagrams (Kalis, Zimmermann 1988). Transitions between the early and middle Neolithic in general are neither simple nor universal. We can also note the situation in the north of Franche-Comté, where the Danubian sequence starts with Roessen (Denaire 2009), while in Lower Saxony, Roessen succeeded Stichbandkeramik (Lönne 2003).

Other concepts in the Strien approach
Finally, we note, but for discussion elsewhere, other facets underpinning Strien’s schema (2017; 2018) for the character and development of the earliest LBK. These include the concept of house generations, an assertion of considerable input by local hunter-gatherers into the formation of the earliest LBK, and the existence, right from the beginning, of marked regionalisation in the distribution of preferred decorative motifs on earliest LBK pottery. All these, which are part and parcel of the Strien schema, could in their turn be disputed, but that would take more space than we have available here. We simply note the many critiques of the Hofplatz model from which the house generation concept is derived (Birkenhagen 2003; Rück 2007; 2013; Lefranc, Denaire in press); though the notion of human generations is useful (references in Whittle 2018), house generations identified only by study of ceramic motifs seem to us at present a hazardous chronometer for tracking change across wide regions. Likewise, the new aDNA evidence already referred to (Lipson et al. 2017) appears to reduce very considerably the likely input of hunter-gatherer populations into the formation and spread of the LBK. Finally, there are also questions about the extent of regionalisation in the earliest LBK (see also Cladders 2018), and we have already noted our classificatory differences with Strien, especially with regard to the proposed formative LBK.

Conclusions
Clearly we consider that virtually all the alleged “discrepancies” between archaeological and 14C-based chronologies postulated by Strien (2017) can be questioned when all the strands of evidence are subject to the detailed and careful scrutiny that is required – constantly and without exception – if we are to come closer to revealing the prehistoric reality that should be the aim of our research. We challenge the apparent opposition of archaeological evidence and 14C dating, considering them rather to be complementary sources of information that, when explicitly combined using a rigorous statistical methodology, together can forge a more reliable understanding of past lives.

We welcome the rigorous questioning of our taken-for-granted, but suggest that this rethinking has to be applied not only to radiocarbon dating, but also to the varied forms of archaeological evidence at our disposal. Dating bone collagen, particularly on the seasonally wet and acidic sands and gravel terraces of the Rhine valley, is undoubtedly challenging. But the accurate dating of such samples has been a major focus of radiocarbon research for the past generation, and it is essential that the existing corpus of dates be interpreted, and new dates obtained, with the benefit of the methodological insights that have been gained. As high-resolution radiocarbon calibration (e.g., Pearson et al. 2018) and high-precision radiocarbon dating by AMS (e.g., Wacker et al. 2010) become available to archaeologists, accuracy must never be assumed, but must always be rigorously tested and evaluated.
Strien (2017.278) is certainly correct in suggesting that cross-checking radiocarbon results on contemporaneous samples of different materials is an effective strategy for testing scientific accuracy, although this is only valid where the taphonomic security of each sample is equivalent. At present, articulating and refitting bones are the most archaeologically secure type of material available to most archaeologists (although we note recent developments in obtaining accurate radiocarbon dates on absorbed fatty acids from pottery (Casanova et al. accepted), which may mean that in future, refitting groups of sherds will provide samples that are as taphonomically secure). But this is only one strategy for assessing the accuracy of radiocarbon dates. We also have at our disposal the routine use of appropriate standard materials in laboratories; interlaboratory replication; the rigorous archaeological, scientific, and statistical identification of misfits and outliers in groups of related dates; formal statistical evaluation of the compatibility of different strands of evidence in a model; and comparison with dendrochronology. All these methods are of value, and we must employ them routinely. Again, we agree with Strien (2017.278) when he asserts that “financial constraints concerning the number of dates are not an excuse for methodological deficits”.

It is equally important to be clear about the taken-for-granted of the chronotypological approach. At its best, this is based on expert, detailed knowledge of the material and its associations, and a related ability to identify securely closed contexts reliably. These virtues can be seen in Strien’s wider study of the earliest LBK (Strien 2018; see also critique in Cladders 2018), and they are behind numerous successful correspondence analyses of important assemblages. But this approach also tends, other things being equal, to assume continuity of use in the material of a given tradition, and to present the results of analysis in such a fashion as to gloss over any possible disruptions or hiatuses (see also critique in Shennan, Wilkinson 2001; Fehl 2015). In that, it is not alone, since many prehistorians of all shapes and sizes have thus far tended to work, often un-thinking, with default perspectives of slow change and continuity (Bayliss et al. 2007; 2016; Whittle 2018). That is often to work away from the nuanced and diverse trajectories of change which we should be attempting to capture in our narratives.

Imagine, however, how much more robust and more effective still our collective efforts could become if the strengths of the various approaches reviewed in this paper were to be applied more regularly and more systematically. In that happy land – perhaps not yet paradise – detailed sequences could be constructed by combining all the strands of information. Short-life samples of known taphonomy, be they animal bones in anatomical order or deposits of cereals, say, in association with diagnostic material culture from assemblages themselves closely ordered by typology or correspondence analysis, or in future the relevant decorated pottery itself, can provide the radiocarbon measurements to input into formal models, along with prior information in the form of detailed archaeological knowledge of context, stratigraphy, and typoarchaeological sequence; the resulting model outputs can then form the basis of interpretation and detailed narrative. That is a Grail worth seeking.

The dating and modelling reported in Jakucs et al. (2016), Oross et al. (2016), Tasić et al. (2016a, 2016b) and Denaire et al. (2017) were carried out within the project The Times of Their Lives, funded by an Advanced Investigator Grant (295412; 2012–2017) from the European Research Council, and led by Alasdair Whittle and Alex Bayliss. The writing of this paper has been supported by a Small Grant to Alasdair Whittle from The British Academy. Grateful thanks are due to both funders. We also thank Kirsty Harding for help with the figures, Dušan Borić, Richard Bradley and Johannes Müller for constructive criticism of a first draft; and Cardiff University for administrative support.
References


Casanova E., Knowles T. D. J., Williams C., Crump M. P., and Evershed R. P. accepted. Practical considerations in high-precision compound-specific radiocarbon dating: eliminating the effects of solvent and sample cross-conta-
minution on accuracy and precision. *Analytical Chemistry*.


back to contents