

Concepts of probability in radiocarbon analysis

Bernhard Weninger¹, Kevan Edinborough², Lee Clare¹ and Olaf Jöris³

¹ Universität zu Köln, Institut für Ur- und Frühgeschichte, Radiocarbon Laboratory, Köln, DE
b.weninger@uni-koeln.de

² Institute of Archaeology, University College London, UK

³ Römisch-Germanisches Zentralmuseum Mainz, Forschungsbereich Altsteinzeit, Neuwied, DE

ABSTRACT – *In this paper we explore the meaning of the word probability, not in general terms, but restricted to the field of radiocarbon dating, where it has the meaning of ‘dating probability assigned to calibrated ¹⁴C-ages’. The intention of our study is to improve our understanding of certain properties of radiocarbon dates, which – although mathematically abstract – are fundamental both for the construction of age models in prehistoric archaeology, as well as for an adequate interpretation of their reliability.*

IZVLEČEK – *V članku raziskujemo pomen besede verjetnost, ne na splošno, temveč omejeno na področje radiokarbonskih datacij, kjer ima beseda pomen ‘verjetnost datiranja dodeljena kalibrirani ¹⁴C starosti’. Namen naše študije je izboljšati naše razumevanje določenih lastnosti radiokarbonskih datumov, ki – čeprav so matematično abstraktni – so temeljnega pomena tako za gradnjo modelov starosti v prazgodovinski arheologiji kot tudi za ustrezne razlage njihove zanesljivosti.*

KEY WORDS – *radiocarbon calibration; Bayesian inference; noncommutative algebra; noncommutative probability; chronology*

Introduction

We begin with a remark attributed to the philosopher and mathematician Bertrand Russell in which he states, clearly not just anecdotally, “*Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means*“. We take this citation from the book ‘Paradoxes in Probability Theory and Mathematical Statistics’, in which the mathematician Gábor Székely (1990) discusses some of the many curiosities that may result from the invalid application of statistical theory. Of immediate relevance to our topic is the date attributed to this anecdote. It appears to have been formulated in the year 1929. If assigned to Russell and correctly dated, albeit that Székely (1990) expresses some doubt concerning both these points, Bertrand Russell would have made this comment some four years prior to the first (German language)

publication of Andrei Nikolaevitch Kolmogorov’s probability theory (Kolmogorov 1933). After its translation into English (Kolmogorov 1956), mathematicians were soon to accept Kolmogorov’s theory, although there were exceptions, including the philosopher of science Karl Popper, who was always dissatisfied with the manner in which Kolmogorov presupposed its foundation in Boolean algebra (Popper 1934; 1959; 1976). In the theory of Kolmogorov, which has three basic axioms, the concept of probability is introduced in Axiom I as a non-negative real number. In Axiom II, this number is limited to a minimum value 0 and maximum value 1. Axiom III defines the mathematical operations (e.g., addition, multiplication) that can be applied to given probabilities in order to produce new probabilities. This is a simplified version of Kolmogorov’s theory,

but one which in our judgement adequately mirrors the manner in which the concept of probability is introduced in the majority of school textbooks today.

It is perhaps of little surprise that statistics and related fields are often experienced as boring and dull. Indeed, and following some three hundred years of advanced and often quite controversial mathematical, philosophical, and even religious discussion, today's widely accepted definition of probability is slightly disappointing. Probability is a number, no more, no less. Some may stress that probabilities are not just any numbers, but rather special, random numbers. Others may emphasise that it is only we, as humans, who have random experiences with numbers, and that numbers themselves cannot support such experiences, in which case the discussion becomes more lively. The problems we address in the present paper are not, however, related to such discourse. We do not participate in the discussion – however far-reaching – as to whether a subjectivist, objectivist, frequentist, or even Bayesian interpretation of probability is the most preferable. Instead, we are quite happy with the notion that probability is nothing more than a number (with a value less than 1) and nothing less than a number (with a value greater than 0), *i.e.*, $0 \leq p \leq 1$.

The noncommutative properties of radiocarbon ages

In the course of the following paper, we assemble some observations to illustrate why, in our judgement, present methods of ^{14}C -analysis do not provide a mathematically sufficient (in the sense of complete) description of the properties of calibrated ^{14}C -ages. We focus here on those properties of ^{14}C -ages that are relevant to the construction of ^{14}C -ba-

sed archaeological chronologies. Unfortunately, the mathematical properties of archaeological ^{14}C -data can be quite subtle as well as misleading, even to the extent of being deceitful, and a clear understanding of these properties is difficult to achieve. Fortunately, the mathematical background to these subtleties can at least be introduced in simple terms, and this is the focus of the following paragraphs.

Generally speaking, mathematical operations are termed *commutative* when study variables can be combined in any order without changing the result. This means that when applied to the addition operation (+) the following equation must be valid: $a + b = b + a$. The same applies to the multiplication operator (x) with $a \times b = b \times a$. We easily confirm that both the operations of addition and multiplication are commutative, at least for the following two integer numbers $a = 2$ and $b = 3$. The test for the commutativity of addition is: $2 + 3 = 3 + 2 = 5$ (confirmed) and for multiplication: $2 \times 3 = 3 \times 2 = 6$ (confirmed).

In mathematics, an operation is called *commutative* if changing the order of the operands does not change the end result. Conversely, *noncommutative* variables are known as *ordered*. Let us now test whether radiocarbon calibration, when represented as a mathematical operator, can be deemed commutative. Note that in the following, we run the commutativity test for numbers, although these are, of course, in place of probabilities (more precisely, probability densities, since probabilities are assigned to intervals).

To simplify the test, we take as an elementary frequency distribution only two numbers on the calendric time-scale, which we call (samples) a and b. Fol-

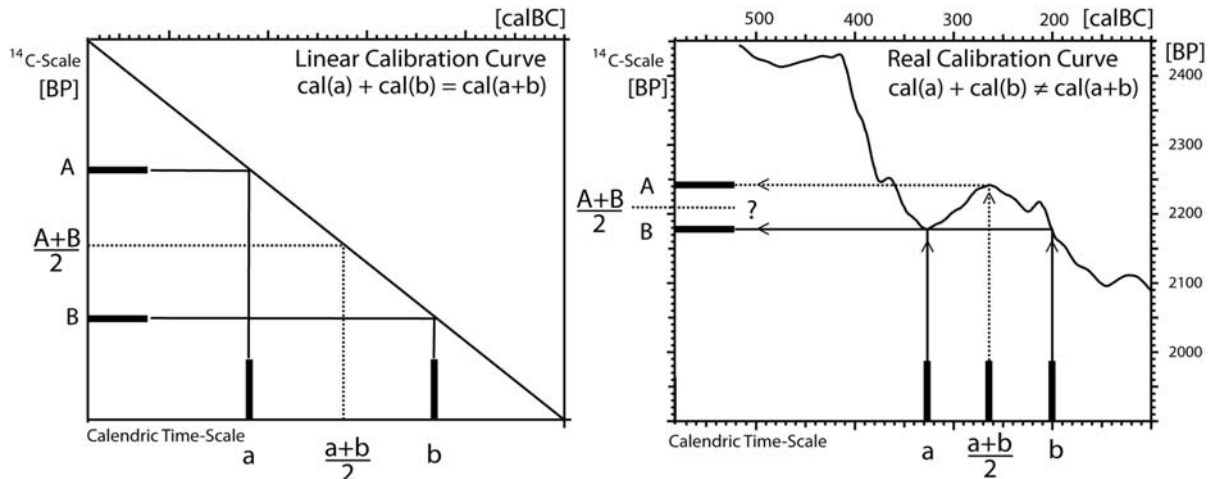


Fig. 1. Commutativity test for the radiocarbon calibration operator.

lowing idealised (error-free) ^{14}C -measurement, samples a and b provide ^{14}C -ages A and B . Assuming that A and B have been correctly obtained, then the initially unknown calendric ages (a and b) can be reconstructed by comparing A and B with previously measured ^{14}C -values for known-age samples taken from the tree-ring ^{14}C -age calibration data-base. The commutativity test for the radiocarbon calibration operator as illustrated in Figure 1.

We test the commutativity of the calibration operator cal in terms of the addition of dates/samples on the two scales (^{14}C - and calendric) by analysing Equation 1:

$$\frac{A + B}{2} = \frac{cal(a) + cal(b)}{2} = \frac{cal(a + b)}{2} \quad (1)$$

This equation may be read as follows: when the measured ^{14}C -age A (for sample a) and the measured ^{14}C -age B (for sample b) are added on the ^{14}C -scale and divided by 2, we obtain the ^{14}C -scale value $(A+B)/2$. This represents the average ^{14}C -age of the two samples. Similarly, the average sample age is calculated on the calendric time scale, where it is defined as $(a+b)/2$. It should be stressed that, due to inherent difficulties concerning the graphic representation of the problem at hand, both in Figure 1 and in Equation 1 the operator cal is in actual fact the inverse calibration.

As shown in Figure 1, for two differently shaped calibration curves, the calibration operator has the property of being commutative *only* under the condition that there is a linear relation between the ^{14}C -scale and the calendric time-scale. Indeed, for the linear calibration curve (Fig. 1, left) identical results are achieved irrespective of whether we first measure the ^{14}C -ages for samples a and b and calculate their average ^{14}C -age on the ^{14}C -scale, or put the carbon of both samples together and measure the ^{14}C -age of the combined sample. For the linear calibration curve the calibration operation is commutative. However, as shown in Figure 1 (right), in the realistic case that the calibration curve is not linear, but wiggly, results obtained by averaging are dependent upon whichever of the two scales the operation is first performed. We abbreviate this by stating that the calibration operator is *noncommutative* in respect to the *addition* of dates and samples on their respective scales.

But why should we be interested in ordering properties of the calibration operator? In archaeology, we are interested in (1) the transfer of ^{14}C -ages from the ^{14}C -scale to the calendric time-scale and (2), in

a second scaling direction, the transfer of sample ages from the calendric time-scale to the ^{14}C -scale. Only the first direction is known *sensu stricto* as ^{14}C -age calibration. However, in archaeological modelling studies, the second operational scaling direction is also of importance, *e.g.*, when the aim is to reconstruct the (theoretical) frequency distribution of samples from the given (measured) distribution of ^{14}C -ages.

Let us further evaluate the commutativity of the calibration operator, but now in terms of multiplication. The multiplication operation is of particular importance in archaeological ^{14}C -analysis, *e.g.*, in Bayesian sequencing methods, when the aim is to constrain the range of calendric ages initially obtained for a larger set of individual ^{14}C -ages. This is accomplished by a combination of the given ^{14}C -data with external (archaeological) information, such as the known or assumed relative-age position of the dated samples. Depending on many details of the specific Bayesian application, the combination of ^{14}C -radiometric and archaeological information inevitably requires the formulation of a mathematical equation, often a complex undertaking. However, a common feature of all such equations is that they contain sums and products of (measured) ^{14}C -scaled probabilities, and of associated (measured or assumed) calendric-scale probability distributions. An example of how complex such an equation can be is illustrated by a case study taken from Christopher Bronk Ramsey (2009). A typical application of the Bayesian sequencing method, in this case applied to a set of ^{14}C -ages from a multiple phase archaeological site, is shown in Equation 2.

$$p(\mathbf{t}) \propto P_H(t_b, t_c) \prod_i \frac{P_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{P_H(t_c, t_j, t_d)}{(t_d - t_c)} \quad (2)$$

We do not provide a description of variables used in this equation, since these are given in detail by Bronk Ramsey (2009, 348). We are interested in the mathematical syntax. As noted above, in Kolmogorov's theory, probabilities are defined as variables with values in the range $0 \leq p \leq 1$. Applying such notions to the above equation, where the different symbols p represent different probabilities, and where the symbol Π is used to abbreviate their multiplication, it becomes evident that Equation 2 contains not only products of probabilities, but – more complicatedly – actually contains products of these products. Nevertheless, and irrespective of the overall complexity of these calculations and whatever the result could mean in archaeological terms, ultimately the resulting overall probability must be a

number in the range $0 \leq p \leq 1$. What is the meaning of the proportional symbol ‘ \propto ’? Clearly, it connects the left and the right side of the equation. Normally, in mathematics the symbol ‘ \propto ’ is used as a dimensionless scaling number (and not as a function), although is this the case here? In Equation 2, the symbol ‘ \propto ’ is apparently used to define a secondary linear scaling operation to be applied at the end of the task. But why should this final scaling operation be necessary? Is there, perhaps, something missing from the equation? As a result of applying the axioms of probability theory, we would usually expect the equation (or algorithm) to be complete in terms of all operations, including scaling (e.g., graphics \rightarrow paper size) and – in particular – in terms of the necessary normalisation of probabilities ($0 \leq p \leq 1$). We will return to this question later, but note that the formulation of Equation 2 using this proportional symbol ‘ \propto ’ differs from Bayes’ theorem where it does not appear (see below).

Clearly, when developing Bayesian sequencing models that include integral-products for probabilities as illustrated by the above equation, great care must be taken that results are not dependent on the order in which the multiplication is performed. To ensure this, the following mathematical equation must be valid:

$$[p_1, p_2] = p_1 \cdot p_2 - p_2 \cdot p_1 = 0 \quad (3)$$

This equation, using Paul Diracs’ commutator notation, can be read as follows: *We call two probabilities p_1 and p_2 commutative when the result of their multiplication is independent of the order in which the multiplication is performed.*

An example would be: $[0.4, 0.8] = 0.4 \times 0.8 - 0.8 \times 0.4 = 0.32 - 0.32 = 0.00$

Since ^{14}C -measurements are defined as rational numbers on the ^{14}C -scale, it is possible – for example, in radiocarbon inter-laboratory exercises – to define ^{14}C -scale probabilities in terms of a Boolean (commutative) event algebra. However, in light of the above observations (Fig. 1) this should not be generalised. To be clear, the problems of commutativity addressed in this paper only arise when the two scales are connected, *i.e.*, as is required for radiocarbon calibration. In light of these observations, we conclude that the calibration operator is noncommutative, both in respect to addition and to multiplication.

Although seemingly rather elementary, the observation that radiocarbon calibration is noncommutative

will resonate as music in the ears of mathematicians and physicists, who – we expect – will be immediately reminded of topics such as quantum mechanics, non-linear geometry, uncertainty principle, wave-particle complementarity, Hilbert Space, C^* -algebra, Lie-algebra and non-classical probability theory.

Looking back, these mathematical properties of the calibration operator indeed resonated, if only as a warning bell, in the ears of the first generation of ^{14}C -calibration software-developers. This we deduce from comments by Minze Stuiver and Paula Reimer (1989), Mieczysław Pazdur and Danuta Michczynska (1989), and Bernhard Weninger (1986), all of whom emphasised the unusual character of the mathematical problem to be solved: *“There are some mathematical pitfalls to be avoided ... We conclude that distortion of [age-calibrated] histograms is unavoidable, even with the most precise mathematical procedure and high-precision ^{14}C dating (Stuiver, Reimer 1989.818, 823).*

The specific mathematical pitfalls in radiocarbon calibration addressed here by Stuiver and Reimer (1989) pertain to the large number of clearly unacceptable oscillations that appear in the calibrated distribution when the calibration is performed from the perspective of the ^{14}C -scale (Fig. 2, right). Fortunately, these oscillations can be avoided by a change in perspective, in which the calibration is performed from the viewing direction of the calendric-time scale (Fig. 2, left).

As illustrated in Figure 3 (right), the technical problem is to see if a horizontal line drawn parallel to the calendric time scale during calibration hits a point on the calibration curve or not. This problem, *i.e.* whether a (2-dimensional) line can actually hit a (1-dimensional) point, or not, is a long-standing problem in the history of geometry.

Yet, this is not the end of the story, as can be recognised from the following comments on this method by Mieczysław Pazdur and Danuta Michczynska (1989.831), who state in dismay: *“The resulting [calibrated] probability distribution in most cases significantly differs from the initial [^{14}C -scale] Gaussian distribution and general rules for simple presentations of calibration output cannot be formulated. Moreover, because of the same reason, the concepts that are widely used and familiar to non-experts in statistics and probability (e.g., mean value, median, confidence interval) begin to lose their seemingly unshakable credibility”.*

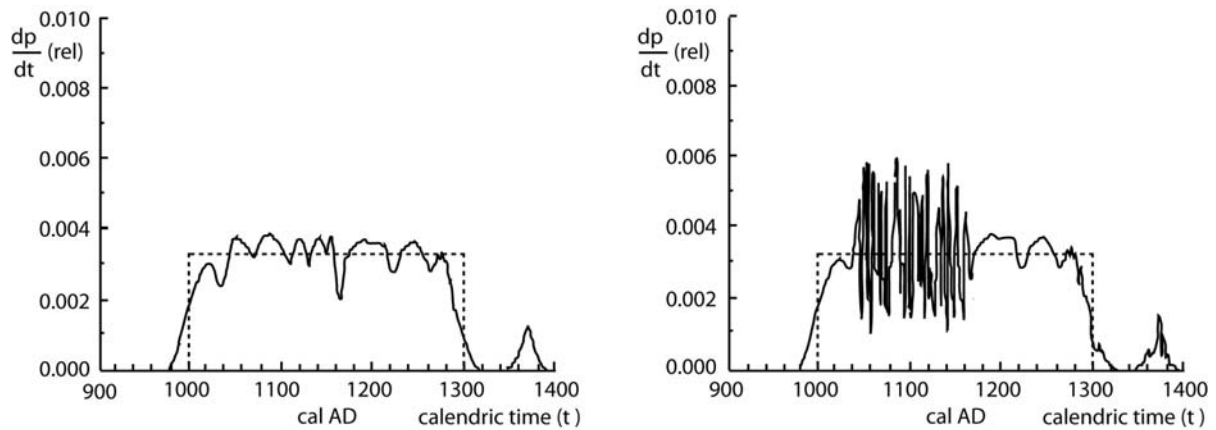


Fig. 2. (Left) ^{14}C -age calibration a ^{14}C -scaled Gaussian probability from the perspective of the calendric time-scale. (Right) ^{14}C -age calibration of the same Gaussian from the perspective of the ^{14}C -scale. Note the existence of unacceptable oscillations in the calendric-scale distribution, when calibration is performed from the perspective of the ^{14}C -scale. Redrawn (with changes in scaling) from Stuiver and Reimer (1989).

To conclude these opening comments on the mathematical problems encountered in ^{14}C -calibration, here is the explanation for the observed effects provided by Bernhard Weninger (1986.27): “A graphic representation of calibrated dates based on Euclidian geometry is not possible. Any method of mapping calibrated dates necessitates construction of a non-linear picture of dating probability”.

In the following years, the approach by which ^{14}C -data are inversely calibrated from the perspective of the calendric time-scale was quickly taken up by the Radiocarbon Community. This approach, since termed *probabilistic* or *Bayesian*, is schematically illustrated in Figure 4. The caption to this figure contains some comments as to the existence of a remaining, and rather awkward, normalisation problem.

Initially, in the late 1980s and early 1990s, only technical reasons were put forward in support of this calibration approach. Clearly, there was a need for its more general foundation in probability theory. This foundation was soon to be identified with the Theorem of Bayes (Pazdur, Michczynska 1898; Michczynska et al. 1990; Niklaus 1993; Dehling, van der Plicht 1993). In retrospect, this decision possibly rests on the simple fact that Bayes’ Theorem is described in much detail in text-books on classical statistics (*cf.* below). Even today, it is taken for granted that Bayesian Theory is applicable, without amendment, to the calibration of radiocarbon ^{14}C -ages. For example, according to Bayliss (2009), Blockley and Housley (2009), Bronk Ramsey (2009), the Bayesian dating methodology provides a coherent framework by which essentially any kind of ^{14}C -

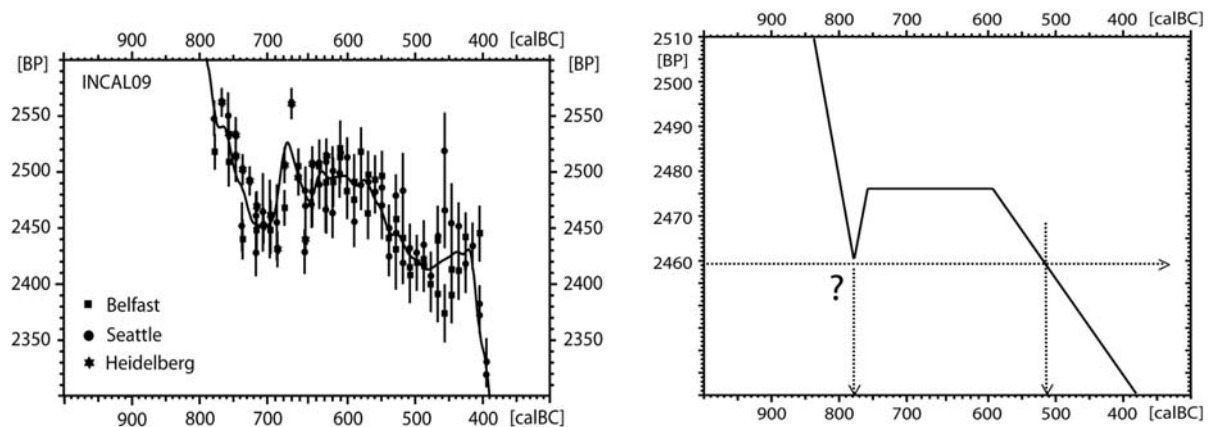


Fig. 3. (Left) Zoom into the ^{14}C -age calibration curve INTCAL09 (Reimer et al. 2009) showing underlying raw data of the High-Precision Laboratories Belfast, Seattle and Heidelberg. (Right) Hypothetical calibration curve with downward spike followed by an extended plateau. Note the difficulty (indicated by the question mark) in providing a unique reading for the ^{14}C -scale value of 2460 BP at the lower end of the spike, but which only exists from the perspective of the ^{14}C -scale. The horizontal dotted line with arrow indicates the reading direction.

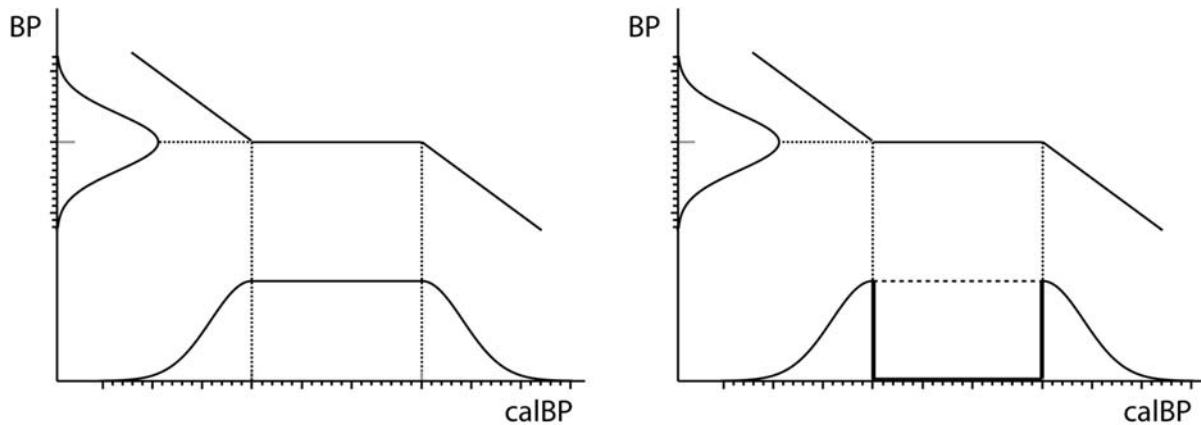


Fig. 4. Schematic application of Bayes' Theorem to the calibration of a Gaussian-shaped ^{14}C -age distribution. For a calibration curve with extended plateau (e.g., 800–400 calBC) there are two (extreme-case) possibilities of transferring a Gaussian shaped ^{14}C -probability from the ^{14}C -scale to the calendric time-scale. (Right) The input Gaussian is divided into two halves in an effort to maintain Kolmogorov's definition of probability as number $0 \leq p \leq 1$. (Left) The plateau region (800–400 calBC) is assigned additional probability. (Right) The resulting archaeological chronology contains artificial gaps. (Left) The archaeological chronology contains artificial enhancement. In both cases, a secondary correction of the chronology is required, the correct application of which is impossible since the true sample age remains unknown. Graph redrawn from Weninger (1986, Fig. 11).

analysis can be performed. Quite unanimously, these authors postulate that the future of ^{14}C -analysis will be a continuation of existing Bayesian calibration concepts. Further to this point, Bronk Ramsey (2009) notes that although a high degree of refinement is still possible, Bayesian analysis of ^{14}C -dates today is a mature methodology sufficiently flexible for adaptation to future needs. We will return to these points, later.

Terminology

Before continuing, our study at this point requires some terminological clarification. In the mathematical sciences and perhaps most clearly in modern physics, the meaning of the term *classical* is quite specific, but it differs significantly from the use of this term by contemporary Radiocarbon scientists. To clarify this issue: in this paper, we use the word *classical* in the sense it is used in quantum physics. In quantum physics, and related mathematical studies, the term *classical* is used to describe the striking dichotomy between the purely probabilistic (*classical*) physical laws that govern (or appear to govern) the macroscopic world, in contrast to the curiously unexpected *non-classical* (*quantum mechanical*) properties that emerge when a closer look is taken at the underlying microscopic properties of the same world. Further, when the term *classical* is used in quantum physics, the intention is to provide a contrast with the results of earlier (e.g., 18th and 19th century) physics theory. Let us take as an example

(which will reappear later in our study) some physical measurements made on certain paired and constrained variables such as energy/time and position/momentum. For these variables, as stated in the Heisenberg uncertainty principle, it is not possible to simultaneously measure the present position and momentum of a quantum-mechanical wave-particle, at least not to unlimited precision, since any measurement of the position of the wave-particle will strongly influence its present momentum, and vice versa. In particular, regardless of the ingenuity of the physical device that is constructed in an effort to avoid measurement uncertainty, the uncertainty principle also applies to the future values of these variables. This was shown in the so-called 'EPR-Gedankenexperiment' thought experiment (Einstein, Podolsky and Rosen 1935).

The immediate analogy to EPR, and for radiocarbon dating if only *cum grano salis*, is that it is equally impossible to actually correct ^{14}C -data for atmospheric ^{14}C -variability. This might appear possible, since the term *radiocarbon calibration* is often (misleadingly) referred to as meaning 'correction of atmospheric ^{14}C -variability'. However, in reality, there is really nothing out there that could be corrected: we can neither eliminate the disturbing atmospheric ^{14}C -variations by any method, nor do we want to change nature. What is actually meant by the term *correction* is that we must allow for past fluctuations of atmospheric ^{14}C -levels in the construction of ^{14}C -based age models in archaeology.

The main point in terminology to be made here, however, is that empirical observations exist whose interpretation may be dependent (sometimes unexpectedly) on the order in which they were made. In physics, as is well-known, the existence of noncommutative variables was first discovered in the early 20th century. However, even before Heisenberg, the study of noncommutative systems was an important branch of mathematics. With research interest increasing strongly in the 1950s, and as such parallel to the study of Bayesian statistical theory, today there are research departments with buildings inhabited from top to bottom with mathematicians whose research is dedicated to the study of noncommutative systems. However, beyond the fact that many people (subjectively) experience the properties of noncommutative variables as curious, in more objective terms there is nothing unnatural about them, although we need to familiarise ourselves with their properties. We will take a closer look at corresponding traits of ¹⁴C-dates below.

The radiocarbon interferometer

Continuing with such analogies, and as ‘*Gedanken-experiment*’, we now introduce the concept that the calibration system can be interpreted as a hypothetical device which we call a radiocarbon interferometer. As with the many devices developed in physics to study the complementary character of wave-particles, in the following we use the radiocarbon interferometer (RI) to observe what happens when ¹⁴C-dates are age-calibrated under controlled conditions (Fig. 5).

For example, we may be interested in studying what happens on age-calibration of a Gaussian-shaped ¹⁴C-scale probability when the shape of the calibration curve is varied. To this purpose, we can choose a certain value on the ¹⁴C-scale, assign to this value a measuring error, let the Gaussian enter the radiocarbon interferometer, and observe its exit on the calendric time-scale. Such an experiment can be performed either on a macroscopic level by choosing a measuring error that is *large* in relation to the amplitude of the calibration curve wiggles, or else on a microscopic level, by assigning dating errors to the Gaussian that are *small* in relation to the amplitude of the calibration wiggles. All that we require in order to run such experiments is corresponding software. In the present paper, we use software called CalPal (Weninger, Jöris 2008).

Radiocarbon calibration at the macroscopic level

A simple model for what we mean by macroscopic level is provided by a linear calibration curve, in which case the calibration operator is commutative. In general terms, on the macroscopic (linear curve) level, we observe that the input-Gaussian is transformed into an exit-Gaussian, *i.e.*, nothing of much interest happens: the input and output Gaussian have the same shape (Fig. 5, Left). However, this situation changes significantly when dates are entered on the microscopic level: We note both a (slight) bimodal change in the shape of the calibrated Gaussian in comparison to the input ¹⁴C-Gaussian, as well as an emerging problem of how to define a central value for the calendric-scale probability distribution, here indicated as a vertical line (Fig. 5, right).

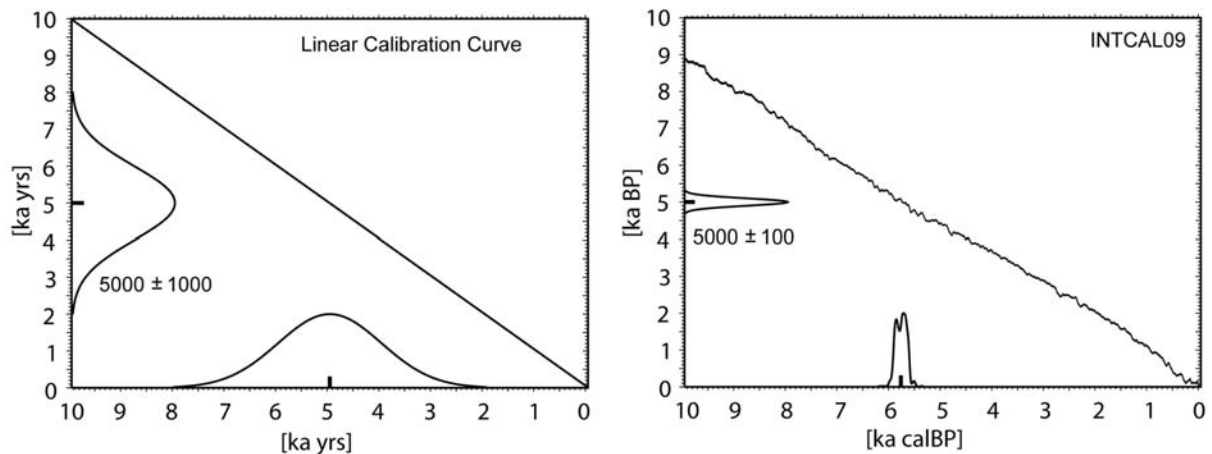


Fig. 5. (Left) Radiocarbon interferometer with linear calibration curve (diagonal line) showing the transfer of a Gaussian-shaped probability distribution from the ¹⁴C-scale (vertical axis) to the calendric time-scale (horizontal axis). (Right) Radiocarbon interferometer with wiggly calibration curve. Note the slight bimodal change of the calibrated Gaussian compared to the input Gaussian, and an emerging (small) offset of the median values (vertical lines) on the calendric time scale.

Radiocarbon calibration at the microscopic level

We define the term microscopic level for cases in which radiocarbon study data show effects (whether in their statistical, geometrical or other properties) that can be attributed to the noncommutative character of the calibration operator. Whether the data show these effects or not is largely a function of dating precision. To this specific point: whereas it is impossible to recognise the atmospheric ^{14}C -variations with single ^{14}C -ages, this proves easier for larger data sets, and easiest when an independent absolute chronology is available for purposes of comparison. In historical terms, the microscopic level that we assign to radiocarbon analysis was already reached around 1970, *e.g.*, with the publication by Hans Suess of the first widely applied tree-ring based ^{14}C -age calibration curve (*Suess 1970*). Using this curve, it was possible for the first time to recognise the existence of secular fluctuations in the global atmospheric ^{14}C -level. The underlying ^{14}C -measurements have standard deviations in the order of $\sigma \sim 80$ BP. However, a later paper by Arie de Jong *et al.* (1979) entitled ‘Confirmation of the Suess wiggles: 3200–3700 BC’ shows that it is no easy task to differentiate conclusively between atmospheric ^{14}C -variability and chance statistical effects in any given data, even for high-precision ^{14}C -measurements ($\sigma < 25$ BP). This is further indicated by the many papers published over the years in which searches to identify the atmospheric ^{14}C -fluctuations in the archaeological ^{14}C -data have been undertaken. The invariable problem lies in the differentiation between atmospheric ^{14}C -fluctuations and chance statistical fluctuations of the study data.

By extension, we might even today expect difficulties in recognising the atmospheric ^{14}C -variations in any given set of real archaeological ^{14}C -ages. For real data, the analytical challenge is to evaluate whether the observed data frequencies are due to the underlying temporal spread of samples, the reconstruction of which would be the analytical goal, or not simply due to a) chance statistical effects, b) the folding properties of the calibration curve, and also – perhaps the most difficult to evaluate – c) the non-commutative properties of the calibration operator as implemented in the specific analytical methodology (software) used.

To begin, a clearly reliable differentiation between atmospheric ^{14}C -fluctuations and chance statistical effects is only possible once the study data have sufficient measuring precision. In physics, the effects as-

sociated with the noncommutativity of certain physical measurements were only recognised once a microscopic level of sensitivity was reached. The same applies to radiocarbon analysis, where the microscopic level is attained only for data with standard deviations in the range of $\sigma \sim 80$ BP, but the smaller the better. Once the data have this precision, then all sorts of interesting effects become apparent on both time-scales (^{14}C - and calendric), not only for larger data sets, but also for single dates.

List of effects caused by the noncommutativity of radiocarbon calibration

Although all the effects caused by the noncommutativity of the calibration operator can be shown for individual ^{14}C -dates, to simplify matters they are introduced in combination in Figure 6. For individual dates, they can be listed as follows:

- dispersal of the exit-Gaussian and its separation into different components on the calendric time-scale;
- lateral shift along the calendric time-scale of the calibrated median;
- dispersal and lateral shift of the area normalised Gaussian also on the ^{14}C -scale;
- separation of calendric-scale confidence intervals into multiple disjunct regions;
- lock-in of numeric-values for confidence intervals (*e.g.*, 95% or 68%) that are used to abbreviate calendric-scale age distributions, also for multiple disjunct intervals;
- the ‘probability values’ assigned to these multiple disjunct intervals seldom sum to 100%.

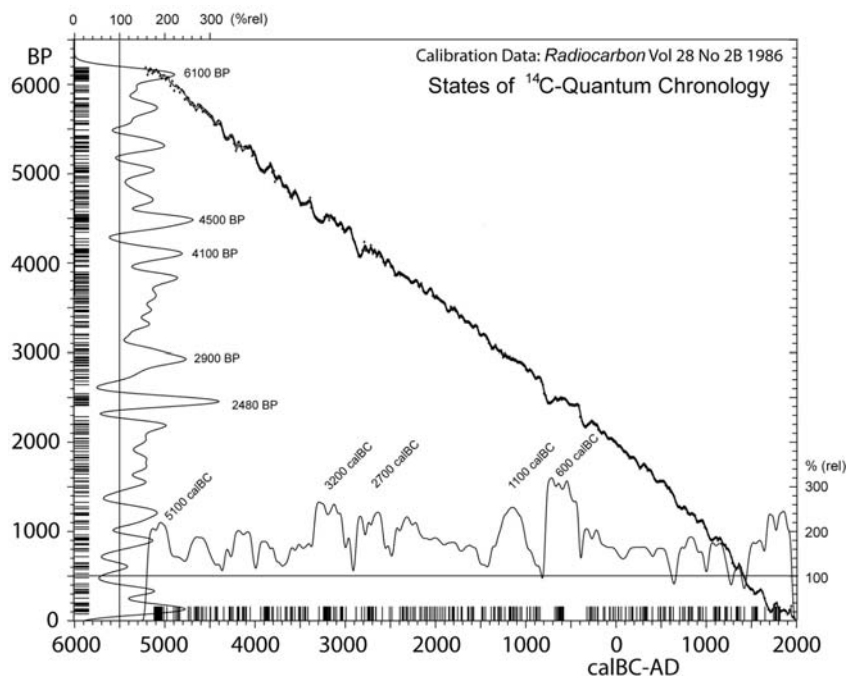
For larger sets of radiocarbon dates, these properties of individual ^{14}C -ages combine to produce the following new effects (Fig. 6):

- clustering of ^{14}C -ages on the ^{14}C -scale;
- clustering of readings on the calendric time-scale;
- attraction of ^{14}C -ages towards predefined intervals on the ^{14}C -scale;
- attraction of calendric readings towards predefined intervals on the calendric scale.

The effects listed above are readily observable both for individual ^{14}C -dates and larger data sets, with no restrictions. In the course of the last three decades, many authors have reported on such properties of ^{14}C -data.

A more complicated issue relates to the question whether it is possible to correct for such effects, as was proposed by Stolk *et al.* (1989; 1994) for the widely used ^{14}C -histogram method (*e.g.*, Geyh 1969;

Fig. 6. The Radiocarbon interferometer, showing the folding properties of the calibration curve for $\sigma = \pm 50$ BP. The initial data entry is a set of $N = 700$ samples artificially placed in decadel increments on the calendric scale, and constructed to provide an error-free uniform calendric-scale sample distribution (not shown, but would be a horizontal line at 100% rel). The ^{14}C -histogram shows a sequence of peaks (e.g., at 6100, 4500, 4100, 2900, 2480 BP) and troughs (not marked) on the ^{14}C -scale. Following the back-calibration of the ^{14}C -histogram, the calibrated age distribution shows a sequence of related peaks (e.g., at 5100, 3200, 2700, 1100, 600 calBC) on the calendric time scale.



on the calendric time scale. Normalisation (First Step): each individual ^{14}C -Gaussian is divided by the ^{14}C -histogram and the individually shape-corrected ^{14}C -Gaussians are added on the ^{14}C -scale. The result is a uniform distribution on the ^{14}C -scale (vertical line at 100% rel). Second step: on the back-calibration, a uniform distribution is obtained in the calendric-scale (horizontal line at 100% rel). Note: this normalisation is hypothetical and only appears to work correctly (cf. text). The small vertical lines represent the median values of the individual dates on both time scales. These lines cluster at peak-positions on both scales. Calculations based on INTCAL86-data (Stuiver, Kra 1986). Redrawn from Weninger (1997).

1971; 1980; Jaguttis-Emden 1977; Breunig, 1987; Weninger 1986; 1997; 2009; Gkiasta et al. 2003; Collard et al. 2010). The underlying idea is to counteract the artificial clustering of ^{14}C -ages, which show up quite clearly in many ^{14}C -histograms, by applying corrections to the histogram shape. This appears possible, since the histogram shape can be calculated to any degree of precision required, for any assumed sample distribution. Hence, it seems necessary only to reduce the histogram amplitude for ^{14}C -values that are known to be artificially enhanced (e.g., calibration curve plateaus) and to enhance the amplitude when the calibration curve is steep (or has few wiggles). In practise, however, this correction introduces an additional distortion of the data frequencies, with an intensity that is typically well beyond that of the initial distortion. This curious problem can be described as follows.

- A uniform age distribution on the calendric time-scale translates into a non-uniform distribution on the ^{14}C -scale. Hence the (entire) ^{14}C -scale histogram can be corrected on that scale, by division by itself, to produce a uniform ^{14}C -distribution. Upon back-calibration, the uniform distribution leads to a uniform distribution on the calendric time-scale *i.e.* the correction appears practicable.

- However, real archaeological data sets contain ^{14}C -ages with admixtures of many different standard deviations. Hence, the necessary shape-correction must be applied to each individual ^{14}C -age. This is technically possible, although it entails considerable number-crunching, since the normalisation function must be calculated and applied to each ^{14}C -Gaussian individually. Alternatively, the shape correction can be performed on the individual age distributions, on the calendric time-scale. But the real problem lies yet deeper: although ^{14}C -histograms may *appear* to be continuous, in fact they show a sequence of discrete events. Hence, in Bayesian analysis we must be cautious in the formulation of prior expectations, since the law of large numbers is not necessarily applicable. If applied, it may produce erroneous results. We further exemplify this particular property of radiocarbon dates in an example below (Fig. 7), and for cases where radiocarbon analysis is taken to what we call the ‘high energy’ extreme.

To conclude the present paragraph, there are a number of analogies between the properties of ^{14}C -ages and corresponding observations made for other non-commutative systems, *e.g.*, in quantum physics. We therefore feel it is legitimate to introduce the notion that radiocarbon dates are ‘quantised’. However,

there is the remaining question as to how the uncertainty principle makes its appearance in the properties of calibrated ^{14}C -data. We address this specific question further below.

Classical Bayesian concepts of radiocarbon analysis

The concepts we are developing in this paper are not necessarily in accord with contemporary notions held by The Radiocarbon Community. As already noted above, according to Bayliss (2009), Blockley and Housley (2009), and Bronk Ramsey (2009), the Bayesian dating methodology already provides a coherent conceptual framework in which essentially any kind of ^{14}C -analysis can be performed. Further to this point, Bronk Ramsey (2009) notes that, although a high degree of refinement is still possible, Bayesian analysis of ^{14}C -dates is already now a mature methodology. This is indicated by the wide variety of existing software that allows for increasingly advanced Bayesian ^{14}C -analysis (software: e.g., BCal; BWigg; CALIB; CalPal; OxCal; cf. Blaauw et al. 2007; Bronk Ramsey, 1994; 1995; 2009; Buck et al. 1996; 1999; Buck 2011; Buck et al. 2008; Christen 1993; Christen et al. 1995; Danzeglocke et al. 2007; Jones, Nicholls 2003; Parnell et al. 2008; Reimer, Reimer 2011; Stuiver, Reimer 1993; van der Plicht 1993; 2011). We can indeed follow all these many authors in their largely unanimous judgement concerning the usefulness, wide applicability, and flexibility of the Bayesian calibration methodology. However, problems remain as to certain mathematical properties of radiocarbon data, e.g., how to adequately define concepts such as *mean value*, *median*, *confidence interval*, and also in view of what the word probability really means when applied to calibrated ^{14}C -ages. Following some twenty years of research in the development of Bayesian concepts of ^{14}C -analysis, we now step back again to the early 1990s to look at the concepts underlying Bayesian ^{14}C -modelling in its very earliest developmental stage.

Although seldom cited, one of the very first applications of the Bayes Theorem to radiocarbon calibration was undertaken by Thomas R. Niklaus, who – in the early 1990s – was compiling his PhD thesis at the Radiocarbon Laboratory of the ETH-Zürich (Eidgenössische Technische Hochschule) in Switzerland. In his PhD, Niklaus (1993) covers technical aspects of the new ^{14}C -AMS-dating method, as well as the development of ^{14}C -age calibration software. With respect to the calibration approach, Niklaus provides

two different but complementary formulations of Bayes Theorem. In a first step, he introduces the basic Bayesian concepts in terms of set theory (cf. Equation 4), and in a second step, he translates these concepts to achieve an integral equation (cf. Equation 5) into which any requested ^{14}C -age distribution can be entered, e.g., Gaussian, in order to achieve a calibrated age distribution.

$$P(X|Y) = \frac{P(XY)}{P(Y)} \quad (4)$$

$$P(XY) = P(X|Y)P(Y) = P(Y|X)P(X)$$

“Die Wahrscheinlichkeit von X unter der zusätzlich erfüllten Bedingung Y wird als bedingte Wahrscheinlichkeit von X unter der Hypothese Y definiert. Die bedingte Wahrscheinlichkeit X lässt sich aus der Wahrscheinlichkeit für die Hypothese Y und der Wahrscheinlichkeit für die Ereignisse X und Y berechnen, wobei direkt daraus das Multiplikationsgesetz {typo: correct would be Gesetz} für bedingte Wahrscheinlichkeiten folgt (Eadie et al. 1971).” (Niklaus 1993:68).

$$f(y|x) = \frac{f(x) f(x|y)}{\int f(x) f(x|y) dx} \quad (5)$$

“Entsprechend der bedingten Wahrscheinlichkeit lässt sich eine bedingte Wahrscheinlichkeitsdichte definieren. Für die bedingten Dichten gilt das folgende Bayes'sche Theorem, welches den Zusammenhang zwischen der Wahrscheinlichkeitsdichte für X unter der Hypothese Y und der entsprechenden bedingten Dichte für Y liefert.” (Niklaus 1993: 68).

Again we are interested in mathematical syntax, not in probability semantics. Hence, we cite here the original German translation of Bayes' Theorem as given by Niklaus (1993), and have therefore not removed the typo of *Multiplikationsgesetz* (correctly, *Multiplikationsgesetz*) i.e. the product rule for probabilities. This typo is understandable, since Gesetz means ‘law’, and gesetzt means ‘to put’. What is important is the introduction of the conditional probabilities $P(X|Y)$ and $P(Y|X)$, in German *bedingte Wahrscheinlichkeiten*. Note, Bayes Theorem is formulated here without requiring the proportional symbol ‘∝’ (cf., above, Equation 2). Interestingly, when applying Bayes' Theorem to provide a mathematical background for radiocarbon calibration, Niklaus (1993) apparently regards it as sufficient to reference a standard statistics text-book (Eadie et al. 1971), ‘Statistical Methods in Experimental Physics’, in which the concept of probability is introduced as a conditional probability. This is quite in accord with

the approach taken in the present paper. On this point, we make specific reference to Popper (1934). After studying in detail the meaning of the term probability in the empirical sciences, Popper concludes – in his *Logik der Forschung* – that probability can always (even in quantum physics) be interpreted as conditional probability.

Due to this restriction of Popper’s research (to the 1930s), we feel it useful to provide a more recent perspective. The question of whether – or not – the probabilistic calibration method can be derived from classical probability theory was first addressed in a seminal paper by Herold Dehling and Hans van der Plicht (1993.244): “*Calibration of radiocarbon dates involves the transformation of a measured ^{14}C age ($BP \pm \sigma$) into a calibrated age distribution (cal AD/BC range). Because of the wiggly nature of the calibration curve, the correct procedure to obtain calibrated age ranges and confidence intervals is not straightforward. Mathematical pitfalls can cause calibration procedures to contradict classical formulas. We show that these ambiguities can be understood in terms of classical and Bayesian approaches to statistical theory.*”

The classical formulas correspond to a uniform prior distribution along the BP axis, the [Bayesian] calibration procedure to a uniform prior distribution along the calendar axis. We argue that the latter is the correct choice, i.e. the [Bayesian] computer programs used for radiocarbon calibration are correct”.

Whereas the first of the above excerpts substantiates the earlier observations made by Stuiver and Reimer (1989), Pazdur and Michczynska (1989), and Weninger (1986), in the second, Dehling and van der Plicht (1993) use the word *classical* to differentiate between two alternative calibration strategies. Whilst the first strategy – calibration from the perspective of the ^{14}C -scale – is referenced to *classical* statistical theory, the second – calibration from the calendric time-scale perspective – is referenced to Bayesian theory. In effect, Dehling and van der Plicht (1993) implement the term *classical* to emphasise the significance of the Bayesian approach in comparison to earlier approaches. An alternative use of the term *classical* is noted in Bronk Ramsey (2009), who states that whereas classical probability theory is aimed at hypothesis testing, the specific idea underlying Bayesian theory is to promote the development of new ideas; however, he often also stresses the non-classical status of Bayesian theory.

In our view, however, it is neither adequate nor important to differentiate between classical and Bayesian probability theory in this manner. Clearly, the classical theory deserves to be called classical. As goes for Bayes’ Theorem, the underlying probability concepts can also be derived from classical theory. In strong contrast to the terminology introduced here by Dehling and van der Plicht (1993), in our view the classical version of probability theory is one that uses commutative variables *i.e.* the classical version represents a rather restricted (special) case of a more general formulation of probability theory which is capable of analysing noncommutative variables. The question is how to provide a mathematically acceptable foundation for noncommutative calibration analysis.

Noncommutative Bayesian concepts of radiocarbon analysis

As indicated by a web-search for topics such as *noncommutative quantum theory* or – more directly – in search of the code-word *noncommutative Bayesian probability*, due to the large number of hits, it immediately becomes clear that mathematical research in this field is widely established and rapidly expanding on a global scale. It is nevertheless difficult to find an elementary introduction to these topics. What we must mainly take into consideration, however, is the existence of some flourishing and even controversial discussions in these fields. In the following paragraph, as exemplification, we provide a brief and comparative review of studies undertaken by Miklos Rédei (Faculty of Natural Sciences, Loránd Eötvös University, Hungary) and by Giovanni Valente (Philosophy Department, University of Maryland, USA). Although both authors address the same question, they come up with entirely different conclusions (Rédei 1992; Valente 2007).

The question, as initially put forward by Rédei (1992) in a paper entitled ‘*When can noncommutative statistical inference be Bayesian?*’ relates to the possibility of extending the concepts of classical Bayesian inference to allow for noncommutative variables. To this aim, Rédei introduces the notion that an abstract rational person (thereafter called ‘agent’) may exist – at least in theory – who is capable of ideally logical thinking. The agent can change his opinion when confronted with some previously unavailable information. We may quantify the agents’ initial degree of belief in event E by assigning to it probability p . Based on some new information, the agent changes his opinion to p' (*i.e.* $p \rightarrow p'$). By ap-

plying Bayes' rule, this change in opinion can be formulated as follows:

$$p'(x) = \frac{p(xE)}{p(E)} = p(x/E) \quad \text{for all } x \text{ in } S \quad (6)$$

To begin, probabilities p and p' are assumed to be defined on a Boolean (commutative) algebra S which contains events E (= Evidence) and other events x . We introduce brackets '()', and call the bracketed events (x) *conclusions*. Initially, the conclusion (x) is supposed to be true (*i.e.*, $p(x) = 1$), but this can change in view of new evidence E , in which case the conclusion (x) may be conditionalised (xE) to provide a revised probability $p(xE)$.

As a '*Gedankenexperiment*' devised to allow for the existence of multiple logically disjunct calibration readings (we may also call E) – let us now allow the agent to review his initial degree of belief (p') in the light of the *same* evidence E , which is put forward a second time. To allow for this 'new' evidence, we further conditionalise (xE) (Equation 5) by including a second E *i.e.* $(xE) \rightarrow (xEE)$ (Eq. 7).

$$p''(x) = \frac{p'(xE)}{p'(E)} = \frac{\frac{p(xEE)}{p(E)}}{\frac{p(xE)}{p(E)}} = p'(x) \quad (7)$$

Since the agent is not provided with different information, on learning a second time of evidence E , he does not have to change his opinion and therefore concludes that $p''(x)$ is identical to $p'(x)$. Although this stability seems to be essential in Bayesian statistical inference, according to Rédei (1992) it does not necessarily apply to the case that the study events are defined for a noncommutative algebraic space (*e.g.*, von Neumann). Given such an algebra, in which the derived probabilities depend on the order in which the events are observed, it may well be the case that $p'' \neq p'$ rather than $p'' = p'$. Having reached this point in his discussion, Rédei (1992) makes reference to the Takesaki Theorem (Takesaki 1972) and concludes there is no satisfactory (*i.e.*, inferentially stable) solution to the problem. In direct analogy, since the events derived from radiocarbon calibration also belong to a noncommutative algebraic space, following Rédei we could now conclude that mathematics does not allow the application of Bayes' Theorem in radiocarbon analysis. As mentioned in the introduction, Popper was never satisfied with the manner in which the Kolmogorow axiomatic theory of probability is founded in (commutative) Boolean algebra. Indeed, in view of the above arguments, his critical notions appear well-founded,

all the more since they apparently apply even to the more general (noncommutative) case.

A significant objection to these notions is provided by Giovanni Valente, however, and his arguments are of immediate relevance to our studies. To begin, Valente (2007) accepts the introduction of a Bayesian agent capable of ideally logical thinking. We note a change in the agents' gender from male (Rédei 1992) to female (Valente 2007), which we adopt in the following. The main point made by Valente (2007) is that – due to the uncertainty principle – in quantum-mechanical experiments the agent cannot be presented with the same evidence, twice. As such, she is never confronted with the conflict scenario, described above, at least not in quantum physics (the topic of Valente's paper). Apparently, even under extreme logical pressure, the agent can always retain her capacity for rational statistical inference. However, this should not be interpreted such that Rédei's arguments are wrong. Simply, according to Valente (2007:840), "*the fact that one cannot have the same evidence twice implies that the stability condition is not applicable in quantum mechanics. Hence ... if the rationality constraint does not apply, one cannot claim its failure*". The analogy to radiocarbon analysis would be that, whenever radiocarbon measurements for the archaeological event-space are replicated, the agent may rationally expect to obtain the same set of ^{14}C -ages, hence – by implication – the same set of calendric time-scale readings.

To sum up, assuming that these arguments (which we have greatly simplified) may be applied to radiocarbon calibration, which we consider reasonable, statistical inference can always (*i.e.* even for noncommutative systems) be formulated within a Bayesian framework.

We raised the question above, of how to obtain a mathematically acceptable foundation for radiocarbon calibration. We conclude this already exists – in Bayes' recipe – and by which we confirm a subset of statements recently made by Bayliss (2009), Blockley and Housley (2009), and Bronk Ramsey (2009) in the 50th Birthday Anniversary edition of the Radiocarbon Journal (Vol. 51, Nr. 1). As it appears, Bayesian calibration analysis is indeed sufficiently flexible to allow for all future refinements in radiocarbon dating, and this includes its necessary reformulation to allow for the noncommutative algebra of radiocarbon calibration. The question remains: are the results of contemporary Bayesian analysis

correct? To be as clear as possible on this point: it is not necessarily the numerical output of available Bayesian calibration software packages that are the immediately critical issue (they can be tested *cf. Steier, Rom 2000*). The crucial question is to identify the (forecast) effects of the uncertainty principle in the properties of calibrated ^{14}C -data.

Technical issues

Pearson theorem

Certainly, there are some technical issues still to be addressed, such as whether or not to assign disjunct confidence intervals to calibrated data. As applies to the calibration of single ^{14}C -ages, we presently prefer to use only the outermost confidence intervals. It is then possible to “*leave the probability distribution within calendrically converted bandwidths to the statisticians*” (Gordon Pearson 1987: 103). This is a visionary statement that we (informally) call Pearson’s Theorem (for an application, *cf.* below). Nonetheless, the rapidly increasing precision and accuracy obtained for ^{14}C -measurements leads, at the latest when reaching high-precision ($\sigma < 25$ BP), to a stabilisation of the often rather ill-defined disjunctness of the multiple calendric-scale intervals. This problem has a practical and a theoretical component. The practical component is that, when applied to lower precision ^{14}C -dates, the obtained list of confidence intervals is often so long as to be unreadable, if not meaningless. For an ideally historical agent (who can anticipate future revisions in the applied calibration curve) the solution would be to cite only the conventional ^{14}C -ages and corresponding laboratory code, both of which are historically stable variables.

The second component refers to the problem that, for calibrated age distributions which are often multimodal, *qua* statistical theory it is not possible to define a meaningful ‘ $\pm 1 \sigma$ ’ (68%-confidence) value. A practical solution is to assign a rectangular probability distribution to the ^{14}C -age, collect a corresponding rectangular distribution on the calendric time-scale, and then apply *Pearson’s Theorem* to its interpretation.

‘Görsdorf’ theorem

For larger sets of data, an analogous notion we call the ‘Görsdorf *Theorem*’ is to imagine that the curve represents the “*envelope over all possible sample distributions*” (*pers. comm., Jochen Görsdorf*). This theorem has recently been generalised by Franz Weninger and other members of the Vienna Environ-

mental Research Accelerator (VERA) to allow for Bayesian Sequencing. Although mathematically rather demanding, by introducing a large (‘infinite’) number of differently shaped prior distributions to work around the intrinsic arbitrariness of Bayesian sequencing based on only one prior, it does seem possible to establish robust Bayesian analysis as a safe sequencing method for ^{14}C -dates (Weninger et al. 2010).

A case study in quantum-theoretical Bayesian calibration

Within the context of the present paper, the following two questions deserve more detailed evaluation: 1) Is it possible to reconstruct the unknown sample distribution by shape-analysis of the corresponding ^{14}C -histogram (Stolk et al. 1989; 1994); and perhaps the most compelling question 2) are there any indications, as forecast by the noncommutative character of the calibration operator, that radiocarbon data show properties that may relate to the quantum-theoretical uncertainty principle? A simultaneous answer can be given to these two questions. This is exemplified in a recently published study by James Steele (2010) in which a direct comparison between the chronological results achieved using different software (CALIB, OxCal, CalPal) for a set of $N = 628$ archaeological ^{14}C -ages (*cf. Buchanan et al. 2011*) is provided (Fig. 7).

When produced with OxCal, (Steele (2010) uses Bronk Ramsey (1995; OxCal version 4.1b3) and CALIB (Steele (2010) uses Stuiver *et al.* (2005; version 5.0), the cumulative data distributions show conspicuous peaks on the calendric time-scale around ~12.9 ka, 11.3 ka, 10.2 ka, 9.5 ka calBP (Fig. 7). Analysing the same data with CalPal, these peaks are virtually non-existent. Steele (2010: 7) comments quite critically on this finding as follows: “*It is immediately obvious that the CalPal output published by Buchanan et al. (2008) has not summed the calibrated probability distributions in the same ways as Calib and OxCal, and that this will have had a significant influence on any visual inference of peaks and troughs in event density. A similar observation about CalPal’s idiosyncratic smoothing algorithm was already made by Culleton (2008).*”

Regarding this second reference, Brendan Culleton (2008:E111) indeed mentions that: “... *CalPal applies a smoothing algorithm to the summed-probability distribution which levels out several sharp*

peaks in the true distribution. The result is an insensitive, low-fidelity population proxy incapable of detecting demographic change.”

However, neither of the two authors provides evidence for their claim that CalPal smooths away some otherwise important peaks in the calibrated age distribution. An alternative interpretation is to take this example as an *experimentum crucis* to localise the exact position within Bayesian calibration methodology where the effects of the quantum-theoretical uncertainty principle become apparent.

Clearly, the four peaks are an artefact of the specific Bayesian algorithms implemented in OxCal and CALIB, but differently in CalPal. This becomes apparent when the age distributions are plotted against the relevant section of the relevant ^{14}C -age calibration curve (INTCAL04). It then becomes visible that the peaks at ~12.9 ka, 11.3 ka, 10.2 ka, 9.5 ka calBP are all positioned along the steepest sections of the ^{14}C -age calibration curve (Fig. 7, upper).

In OxCal and CALIB, a uniform prior is applied to the data frequencies based on the (plausible) assumption that all calendric ages have equal dating probability. Mathematically, the implementation of this prior involves providing an equal-area normalisation to the dates. Whether the normalisation is undertaken on the ^{14}C -scale or on the calendric scale is of little consequence. Both are technically possible, and in both cases the normalisation corresponds to the same assumption, namely that it is possible to simultaneously correct both the shape of the ^{14}C -histogram and the shape of the calibrated data frequency distribution in order to allow for its distortion due to the non-linearity of the calibration curve. In contrast to this intention, what actually happens as a result of frequency normalisation is that the distortive effects are further enhanced. As such, instead of the intended correction, the frequency normalisation actually over-corrects the data to an unacceptable degree.

The question at stake is on which of the two scales – if any – should the corrections be applied? OxCal and CALIB both make use of a uniform prior on the calendric scale, and apply corresponding corrections to the posterior data frequency on the calendric time-scale. In CalPal the underlying Bayesian assumption is that, similarly, both on the calendric-scale and on the ^{14}C -scale, there exists a uniform prior dating probability. However, since this double *a priori* assumption is neither plausible nor validated by archaeological reasoning, CalPal does not apply the resulting correction to the posterior data frequency, *i.e.* the ^{14}C -histogram is simply transferred from the ^{14}C -scale to the calendric time-scale without further correction (by applying Görzsdorf’s Theorem *cf.* above). From a (classical) Bayesian perspective, this approach may appear to be contradic-

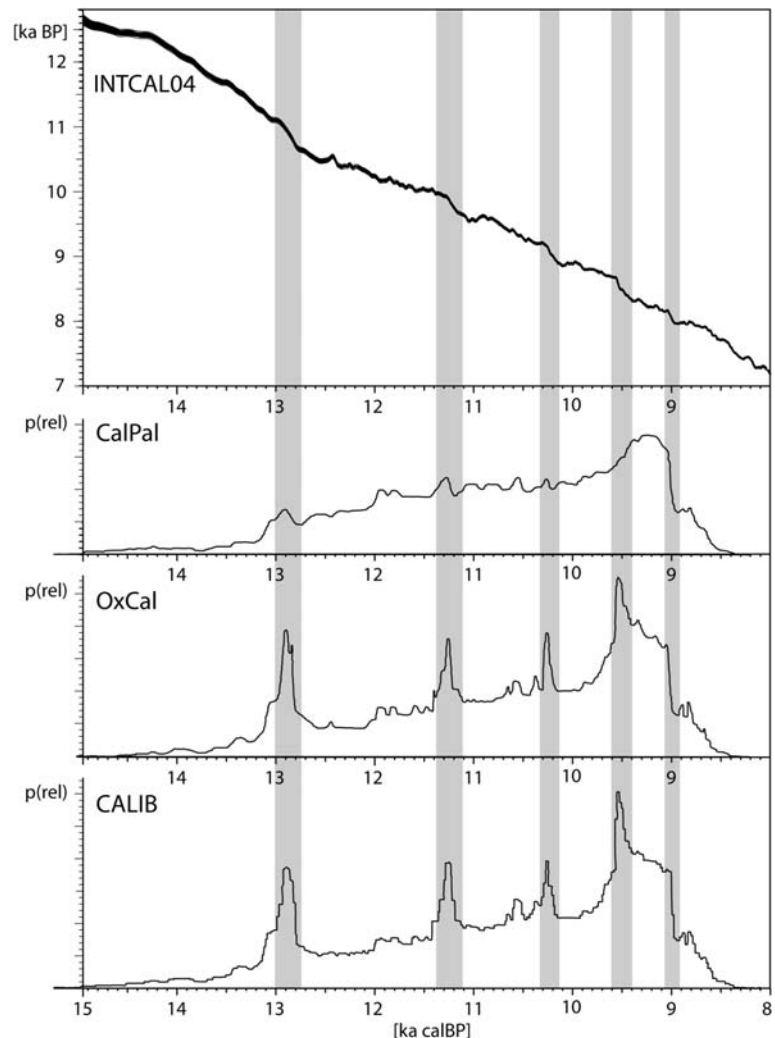


Fig. 7. Cumulative probability distributions, each calculated for the same $N = 628$ ^{14}C -ages using different software (CALIB, OxCal, CalPal). Shaded areas put focus on the correlation of peaks with steep sections of INTCAL04. Graph redrawn from Steele (2010, Fig.5). (Upper) Insertion of INTCAL04 calibration curve (Reimer et al. 2004).

tory. As argued above, however, the ultimate problem is that there is a noncommutative algebraic relation between the two scales.

In mathematical terms, whereas OxCal and Calib are based on a classical Bayesian approach to ^{14}C -calibration, CalPal applies non-classical quantum-theoretical (QT) Bayesian probability concepts. The uncertainty relation well-known in quantum physics thus re-appears, within the framework of QT-calibration, in the manner that a simultaneously correct measurement (or reconstruction) of data probability functions, on the two time-scales, is not possible. Interestingly, in QT-calibration, the uncertainty relation is one- and not two-sided *i.e.* we can calculate the shape of the ^{14}C -scale frequency distribution perfectly for any given set of calendric-scale events, but not the reverse.

Radiocarbon calibration at the sub-microscopic level

In the following (final) paragraph, by taking radiocarbon calibration to the sub-microscopic level, we turn our attention to the future of Bayesian radiocarbon analysis. Above, we have made repeated use of certain analogies between the noncommutative properties of the calibration operator and of corresponding properties of waves and particles in quantum physics. In extension, and by taking these analogies both seriously and one step further, we may now forecast that radiocarbon dates (alias wave-particles) should actually show further effects, but which can only become apparent when the associated ‘energy’ is taken to the extremes. In physics, the energy associated with a wave-particle is related to the frequency of the wave, such that the higher the frequency of the wave, the higher the associated ener-

gy. In particular, whereas at low energies the wave character of elementary particles is the most apparent, the particle character becomes increasingly more apparent at higher energies. Taken to the extreme, when a certain energy limit is reached, the total energy can – and quite often actually does – allow for the production of new particles. Obviously, we are using such analogies to radiocarbon data really only *cum grano salis*, but nevertheless, it is interesting to apply such concepts to radiocarbon dates. The direct analogy to radiocarbon dates would be to associate their energy-content with the precision of the ^{14}C -measurements. With this analogy, what we would expect is that, when increasingly smaller standard deviations are applied to the ^{14}C -scale Gaussian distribution, the larger and more apparent should be the forecast quantum-theoretical effects.

In a further ‘*Gedankenexperiment*’, again using the hypothetical radiocarbon interferometer as an experimental device, let us analyse what happens when calibrating a Gaussian shaped ^{14}C -scale probability distribution, when only the measuring precision of raw data underlying the calibration curve is taken to an extreme. In this case, the calibration curve could well take on a (hypothetical) zigzag shape as illustrated in Figure 8.

Although the calibrated age distributions show strong oscillations, the forecast particle character does not yet become apparent. In a second ‘*Gedankenexperiment*’, let us therefore simultaneously maximise both the measurement precision of the calibration curve and the precision of the archaeological ^{14}C -ages to be calibrated.

The results are now as forecast. As illustrated in Figure 9 for each of the two independently measured

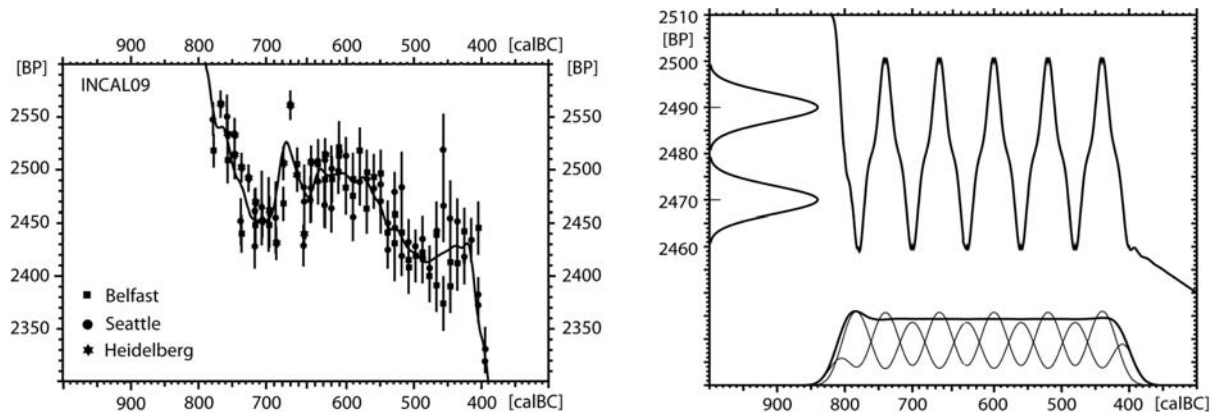


Fig. 8. (Left) Zoom into the presently recommended ^{14}C -calibration curve (INTCAL09) showing underlying raw data of the High-Precision Laboratories Belfast, Seattle and Heidelberg. (Right) Hypothetical zigzag calibration curve.

'high-energy' ^{14}C -dates (Dates A and B), what we observe is that corresponding readings in the calendric time-scale are now clearly separated, *i.e.* they show no temporal overlap. The readings attributed to the two different Dates (A and B) also show no temporal overlap.

In effect, at the 'high energy' extreme – although only single (^{14}C -scale) particles have been entered into the radiocarbon interferometer – the noncommutative properties of the calibration operator (and corresponding folding properties of the calibration curve) are now sufficiently strong to produce some previously non-existent particles 'out of a vacuum' on the calendric time-scale.

We may now re-formulate the question whether – or not – a correction of the ^{14}C -histogram shape (or corresponding shape of the calibrated frequency distribution) to allow for the folding properties of the calibration curve is possible. As already noted above, the correction must be applied to the probability distribution for each individual date. Now that the distribution is reduced to a series of digital 'true-false' (yes-no; 1/0) decisions (Fig. 9), it becomes apparent that the correction – to be applicable – must assign an individual truth-value to each of the alternative readings. Since the readings are mutually exclusive, only one of these values can be 'true'. If this cannot be achieved, the analysis will produce contradictory results, at least if we apply the notion that a proposition is either true or false, and that a third solution does not exist. This would accord with the *tertium-non-datur* of scholastic logic, in which the statement (A and \neg A) always has the truth-value 'false'. However, in radiocarbon calibration, we must allow for a third possibility: the truth-value assigned to any specific calendric age interval may remain unknown. As it appears, therefore, once the 'high-energy' extreme is reached, the analysis is immediately confronted with a Bayesian inference problem (as described above):

- ① Since there is no temporal overlap of the event sequence (Fig. 9), conditional multiplication of the associated probabilities will *always* produce the value $p = 0$.
- ② Since the number of events greatly increases with the number of curve-wiggles, and in particular, faster than we may (perhaps) be able to provide

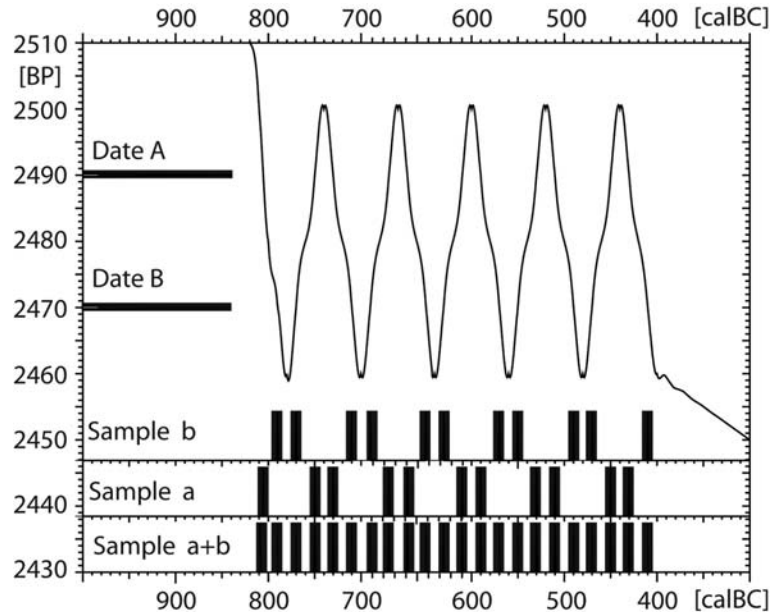


Fig. 9. Radiocarbon calibration at 'high energy'.

additional ^{14}C -ages (or other conditional dating information) from the archaeological stratigraphy, perhaps we may never be able to provide a sufficient number of ^{14}C -ages to catch up with the number of readings. Looking back at what Rédei's agent would have concluded when confronted with the same information twice (or even more often), it appears – maybe even more frustrating (Rédei 1992.2) – that the noncommutative character of the calibration operator is not even stable, but can vary strongly along the calendric-time scale (*i.e.* within the limits of the non-Boolean algebra as given by the tree-ring ^{14}C -age calibration data set). Fortunately, again following Rédei (1992.5), we may disregard this specific problem, since one does not expect mathematical theorems to give insight into the psychological processes of the human mind.

Conclusions

The last '*Gedankenexperiment*' takes us to the very limits of radiocarbon dating. Having arrived at this critical point, we must now emphasise that – even under such extreme analytical conditions – we have no reason to seriously question the applicability of the Bayesian approach to radiocarbon analysis. To be sure, as recently pointed out again by Peter Steier and Werner Rom (2000), there are many applications where the prior information necessary to delimit the number of disjunct readings is known in full detail (*e.g.*, in tree-ring 'wiggles matching'). In such cases, the available information can be transformed into a Bayesian mathematical form which is

capable of providing a closed (although not always unique) mathematical solution to the dating problem (*Bronk Ramsey et al. 2001*). Concerning Bayesian sequencing, the advantage of this method is that it is state-of-the-art; however, it will surely be advantageous to develop graphic methods that allow the user to actually visualise its chronological (quantum) limits. Although a challenging undertaking, this would also provide a solution to the problem that – under certain conditions – Bayesian sequencing is known to optimise the precision of the dating at the expense of its accuracy (*Steier, Rom 2000*).

What archaeologists can do in support of such futuristic efforts is to provide Bayesian sequencing with as much (quantitative) archaeological input as possible, along with correspondingly complete error analysis. Such information may be derived *e.g.*, by careful selection of single-event samples from high-resolution archaeological stratigraphies, by the application of pottery (or other) seriation, by sequencing of samples with well-defined positions in Harris matrices, and – last, but not least – by the development of architectural (*e.g.*, house construction, use, destruction, abandonment) as well as cultural (*e.g.*, demographic) archaeological models. By way of the rule ‘*the higher the requested dating precision, the more samples must be dated to circumvent the simultaneously increasing number of wiggles*’, it may also be necessary to convince funding agencies that such dating efforts are really worthwhile.

Most importantly, however, perhaps we should not overlook the simple fact that it has never been claimed that Bayesian analysis can provide a closed solution to all archaeological applications, under all circumstances. The Bayesian method began as an entirely probabilistic approach some 250 years ago (*Bayes 1763*), was further developed as such by mathematicians to allow for the incorporation of revised probability concepts (*e.g.*, Kolmogorov’s axiomatic foundation of commutative probability theory), and has even survived the many (still running) intellectual revolutions in physics and science-philosophy that resulted from the introduction of non-commutative probabilities.

We have formulated our studies in (hopefully) understandable language, such that radiocarbon scientists and archaeologists alike may be interested in their critical evaluation. If there is need for some clearly formulated single conclusion, it would be as follows: In the past, radiocarbon dating probability was taken to represent a value (number) attributed to each interval of the calendric time-scale. We need not change this notion. The new quantum probability is again a number. However, it is no longer valid to assume that the larger this number, the more probable the dating.

— ACKNOWLEDGEMENTS —

We wish to thank Professor Mihael Budja for encouraging these studies.

∴

REFERENCES

- BAYES T. R. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society* 53: 370–418.
- BAYLISS A. 2009. Rolling Out Revolution; Using Radiocarbon Dating in Archaeology. *Radiocarbon* 51(1): 123–147.
- BLAAUW M., BAKKER R., CHRISTEN J. A., HALL V. A. and VAN DER PLICHT J. 2007. A Bayesian Framework for age modelling of Radiocarbon-dated peat deposits: Case studies from the Netherlands. *Radiocarbon* 49(2): 357–367.
- BLOCKLEY S. E., HOUSLEY R. A. 2009. Calibration Commentary. *Radiocarbon* 51(1): 287–290.
- BREUNIG P. 1987. *14C-Chronologie des vorderasiatischen, südost- u. mitteleuropäischen Neolithikums*. Fundamenta Reihe A, 13. Köln/Wien.
- BRONK RAMSEY C., VAN DER PLICHT J. and WENINGER B. 2001. ‘Wiggle matching’ radiocarbon dates. *Radiocarbon* 43(2A): 381–389.
- BRONK RAMSEY C. 1994. Analysis of chronological information and radiocarbon calibration: the program OxCal. *Archaeological Computing Newsletter* 41: 11–6.
1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2): 425–430.

2009. Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* 51(1): 337–360.
- BUCHANAN B., HAMILTON M., EDINBOROUGH K., O'BRIEN M. J. and COLLARD M. A. 2011. Comment on Steele's (2010) "Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities". *Journal of Archaeological Science* 38(9): 2116–2122, doi: 10.1016/j.jas.2011.02.026
- BUCK C. E. 2011. BCal, an on-line radiocarbon calibration tool: <http://bcal.shef.ac.uk>
- BUCK C. E., CAVANAGH W. G., LITTON C. D. 1996. *Bayesian approach to Interpreting Archaeological Data*. Wiley, Chichester.
- BUCK C. E., CHRISTEN J. A. and JAMES G. N. 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeology* 7. online http://intarch.ac.uk/journal/issue13/christen_index.html
- CHRISTEN J. A. 1993. Bwigg: an internet facility for Bayesian Radiocarbon Wiggle Matching. *Internet Archaeology* 13. online http://intarch.ac.uk/journal/issue13/christen_index.html
- CHRISTEN J. A., LITTON C. D. 1995. A Bayesian approach to wiggle-matching. *Journal of Archaeological Science* 22(6): 719–725.
- COLLARD M., EDINBOROUGH K., SHENNAN S. and THOMAS M. G. 2010. Radiocarbon Evidence indicates that migrants introduced farming to Britain. *Journal of Archaeological Science* 37: 866–870.
- CONARD N. J., 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: 331–371.
- CULLETON B. J. 2008. Crude demographic proxy reveals nothing about Paleoindian population. *Proceedings of the National Academy of Sciences of the United States of America* 105 (50): E111.
- DANZEGLOCKE U., WENINGER B. and JÖRIS O. 2007. On-line Radiocarbon Age Calibration: www.calpal-online.de. online <http://download.calpal.de/calpal-download/>
- DEHLING H., VAN DER PLICHT J. 1993. Statistical Problems in Calibrating Radiocarbon Dates. *Radiocarbon* 35 (1): 239–244.
- DIRAC P. A. M. 1947. *The Principles of Quantum Mechanics*. 3rd edition. Oxford University Press. Oxford.
- DE JONG A. F. M., MOOK W. G. and BECKER B. 1979. Confirmation of the Suess wiggles: 3200–3700 BC. *Nature* 280: 48–49.
- EADIE W. D., DRYARD D., JAMES F. E., ROOS M. and SA-DOULET B. 1971. *Statistical Methods in Experimental Physics*. North Holland Publishing Co. Amsterdam: 17–23.
- EINSTEIN A., PODOLSKY B. and ROSEN N. 1935. Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 47: 777–780, doi:10.1103/PhysRev.47.777.
- FAIRBANKS R. G., MORTLOCK R. A., CHIU T.-C., CAO L., KAPLAN A., GUILDERSON T. P., FAIRBANKS T. W., BLOOM A. L., GROOTES P. M. and NADEAU M.-J. 2005. Radiocarbon calibration curve spanning 10,000 to 50,000 years BP based on paired $^{230}\text{Th}/^{234}\text{U}/^{238}\text{U}$ and ^{14}C dates on pristine corals. *Quaternary Science Reviews* 25: 1781–1796.
- GEYH M. 1969. Versuch einer chronologischen Gliederung des marinen Holozän an der Nordseeküste mit Hilfe der statistischen Auswertung von ^{14}C -Daten. *Zeitschrift der Deutschen Geologischen Gesellschaft* 188(2): 351–360.
1971. Middle and young Holocene sea-level changes as global contemporary events. *Geologiska Föreningens i Stockholm Förhandlingar* 93: 679–692.
1980. Holocene sea-level history: Case study of the statistical evaluation of ^{14}C dates. In M. Stuiver and R. S. Kra (eds.), Proceedings of the 10th International ^{14}C Conference. *Radiocarbon* 22(3): 695–704.
- GKIASTA M., RUSSELL T., SHENNAN S. and STEELE J. 2003. Neolithic Transition in Europe: the radiocarbon record revisited. *Antiquity* 77(295): 45–62.
- JAGUTTIS-EMDEN M. 1977. *Zur Präzision archäologischer Datierungen*. Archaeologica Venatoria 4. Tübingen.
- JONES M., NICHOLLS G. K. 2003. New radiocarbon calibration software. *Radiocarbon* 44(3): 663–674.
- KOLMOGOROV A. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik. 2. Band, Heft 3, Berlin. online <http://www.mathematik.com/Kolmogorov/index.html>
- KOLMOGOROV A. N. 1956. *Foundations of the Theory of Probability*. New York. Chelsea Publishing Company.
- MICHYZYNSKA D. J., PAZDUR M. F. and WALANUS A. 1990. Bayesian approach to probabilistic calibration of radiocarbon ages. In W. G. Mook and H. T. Waterbolk (eds.), *Proceedings of the 2nd International Symposium ^{14}C and Archaeology*. PACT 29. Strasbourg: 69–79.

- NIKLAUS T. R. 1993. Beschleunigermassenspektrometrie von Radioisotopen. Unpublished doctoral dissertation University of Zürich. Zürich. ETH Nr. 10065, doi:10.3929/ethz-a-000697402
- OTTAWAY J. H., OTTAWAY B. 1975. Irregularities in the Dendrochronological Calibration Curve. In T. Watkins (ed.), *Radiocarbon, Calibration and Prehistory*. Edinburgh University Press, Edinburgh: 28–38.
- PARNELL A. C., HASLETT J., ALLEN J. R. M., BUCK C. E. and HUNTLEY B. 2008. A flexible approach to assessing synchronicity of past events using Bayesian reconstructions of sedimentation history. *Quaternary Science Reviews* 27: 1872–1885.
- PAZDUR M. F., MICHCZYNSKA D. J. 1989. Improvement of the procedure for probabilistic calibration of radiocarbon. *Radiocarbon* 31(3): 824–832.
- PEARSON G. W. 1986. Precise calendrical dating of known growth-period samples using a ‘curve fitting’ technique. *Radiocarbon* 28: 292–299.
1987. How to cope with calibration. *Antiquity* 61(231): 98–103.
- POPPER K. R. 1934. *Logik der Forschung*. Julius Springer. Wien.
1976. *Logik der Forschung*. J. C. B. Mohr (Paul Siebeck). Tübingen.
- RÉDEI M. 1992. When can non-commutative statistical inference be Bayesian? *International Studies in the Philosophy of Science* 6: 129–132. <http://phil.elte.hu/~redei/cikkek/bayesi.pdf>
- REIMER P. J., REIMER R. 2011. CALIB. Online <http://calib.qub.ac.uk/calib/>
- REIMER P. J., BAILLIE M. G. L., BARD E., BAYLISS A., BECK J. W., BERTRAND C. J. H., BLACKWELL P. G., BUCK C. E., BURR G. S., CUTLER K. B., DAMON P. E., EDWARDS R. L., FAIRBANKS R. G., FRIEDRICH M., GUILDERSON T. P., HOGG A. G., HUGHEN K. A., KROMER B., MCCORMAC F. G., MANNING S. W., RAMSEY C. B., REIMER R. W., REMMELE S., SOUTHON J. R., STUIVER M., TALAMO S., TAYLOR F. W., VAN DER PLICHT J. and WEYHENMEYER C. E. 2004. IntCal04 terrestrial radiocarbon age calibration, 26–0 ka BP. *Radiocarbon* 46: 1029–1058.
- 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 A. 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: 1111–1150.
- STEIER P., ROM W. 2000. The use of Bayesian statistics for 14C dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42(2): 183–98.
- STEELE J. 2010 Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities. *Journal of Archaeological Science* 37: 2017–2030.
- STOLK A. D., HOGERVORST K. and BERENDSEN H. 1989. Correcting 14C Histograms for the Non-Linearity of the Radiocarbon Time Scale. *Radiocarbon* 31(2): 169–178.
- STOLK A. D., TÖRNQUIST T. E., KILIAN P., BERENDSEN H. J. A. and VAN DER PLICHT J. 1994. Calibration of 14C Histograms: A Comparison of Methods. *Radiocarbon* 36(1): 1–10.
- STUIVER M., KRA R. 1986. Radiocarbon Calibration Issue. Proceedings of the 12th International Radiocarbon Conference, June 24–28, Trondheim, Norway. In M. Stuiver and R. Kra (eds.), *Radiocarbon* 28(2B): 805–1030.
- STUIVER M., REIMER P. J. 1989. Histograms obtained from computerized radiocarbon age calibration. *Radiocarbon* 31(3): 817–823.
1993. Extended 14C data base and revised CALIB 3.0 14C age calibration program. *Radiocarbon* 35(1): 215–230.
- STUIVER M., REIMER P. J. and REIMER R. W. 2005. CALIB5.0. online <http://calib.qub.ac.uk/calib/>
- Suess H. E., 1970. Bristlecone pine calibration of the radiocarbon time-scale 5200 BC to the present. In I. U. Olson (ed.), *Radiocarbon variations and absolute chronology. Proceedings of Nobel symposium, 12th*. John Wiley & Sons, New York: 303–311.
- SZÉKELY G. J. 1990. *Paradoxa. Klassische und neue Überraschungen aus Wahrscheinlichkeitstheorie und mathematischer Statistik*. Verlag Harri Deutsch. Thun und Frankfurt am Main.
- TAKESAKI M. 1972. Conditional Expectations in von Neumann Algebras. *Journal of Functional Analysis* 9: 306.
- VALENTE G. 2007. Is there a stability problem for Bayesian noncommutative probabilities? *Studies in History and Philosophy of Modern Physics* 38: 832–843.

VAN ANDEL T. H. 2005. The ownership of time: approved ^{14}C calibration or freedom of choice? *Antiquity* 79: 944–948.

VAN DER PLICHT T. H. 1993. The Groningen Radiocarbon Calibration Program. *Radiocarbon* 35(1): 231–237.

2011. WinCal25. online <http://www.rug.nl/ees/informatieVoor/cioKlanten?lang=en>

WENINGER B. 1986. High-precision calibration of archaeological radiocarbon dates. *Acta Interdisciplinaria Archaeol* 4: 11–53.

1997. *Studien zur dendrochronologischen Kalibration von archäologischen ^{14}C -Daten*. Habelt Verlag, Frankfurt/M.

WENINGER B., JÖRIS O. 2008. A ^{14}C age calibration curve for the last 60 ka: the Greenland-Hulu U/Th timescale and its impact on understanding the Middle to Upper Paleolithic transition in Western Eurasia. *Journal of Human Evolution* 55: 772–781.

WENINGER B., CLARE L., ROHLING E., BAR-YOSEF O., BOHNER U., BUDJA M., BUNDSCHUH M., FEURDEAN A., GEBEL H-G, JORIS O., LINSTADTER J., MAYEWSKI P., MUHLENBRUCH T., REINGRUBER A., ROLLEFSON G., SCHYLE D., THISSEN L., TODOROVA H. and ZIELHOFER C. 2009. The Impact of Rapid Climate Change on prehistoric societies during the Holocene in the Eastern Mediterranean. In M. Budja (ed.), *16th Neolithic Studies. Documenta Praehistorica* 36: 7–59.

WENINGER F., STEIER P., KUTSCHERA W. and WILD E. M. 2010. Robust Bayesian Analysis, an attempt to improve Bayesian Sequencing. *Radiocarbon* 52(2–3): 962–983.